PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 21/06, 1/68, C12N 15/00, 1/12, 5/00, 5/02

A1

(11) International Publication Number: WO 99/18243

(43) International Publication Date: 15 April 1999 (15.04.99)

(21) International Application Number: PCT/US98/21151

(22) International Filing Date: 6 October 1998 (06.10.98)

(30) Priority Data:

60/061,143	6 October 1997 (06.10.97)	US
60/061,149	6 October 1997 (06.10.97)	US
60/061,159	6 October 1997 (06.10.97)	US
09/004,206	8 January 1998 (08.01.98)	US
09/010,674	22 January 1998 (22.01.98)	US
09/014,347	27 January 1998 (27.01.98)	US

(71) Applicant: MILLENNIUM BIOTHERAPEUTICS, INC. [US/US]; 620 Memorial Drive, Cambridge, MA 02139 (US).

(72) Inventors: PAN, Yang; 68 Hamilton Road #1, Brookline, MA 02446 (US). GEARING, David, P.; 23 Standish Road, Wellesley, MA 02181 (US). McCARTHY, Sean, A.; 62 Commonwealth Avenue #4, Boston, MA 02116 (US).

(74) Agents: MANDRAGOURAS, Amy, E. et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SIGNAL PEPTIDE CONTAINING PROTEINS AND USES THEREFOR

(57) Abstract

Novel LSP-1, PA-I, and TAP-1 polypeptides, proteins, and nucleic acid molecules are disclosed. In addition to isolated, full-length LSP-1, PA-I, and TAP-1 proteins, the invention further provides isolated LSP-1, PA-I, and TAP-1 fusion proteins, antigenic peptides and anti-LSP-1, anti-PA-I, and anti-TAP-1 antibodies. The invention also provides LSP-1, PA-I, and TAP-1 nucleic acid molecules, recombinant expression vectors containing a nucleic acid molecule of the invention, host cells into which the expression vectors have been introduced and non-human transgenic animals in which a LSP-1, PA-I, and TAP-1 gene has been introduced or disrupted. Diagnostic, screening and therapeutic methods utilizing compositions of the invention are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

SIGNAL PEPTIDE CONTAINING PROTEINS AND USES THEREFOR

Background of the Invention

Cells of the immune response characteristically express a variety of cell-surface proteins which are crucial to proper functioning of the immune system. Such proteins include surface immunoglobulins, non-immunoglobulin cell surface antigen receptors, cellular adhesion molecules, as well as other selected phenotypic markers. Many of these cell surface proteins are members of the immunoglobulin (Ig)superfamily of proteins, characterized by the existence of at least one immunoglobulin (Ig) domain. Such proteins function in an variety of immune cell functions ranging from immune cell development and differentiation, antigen recognition, antibody production, cellular signal transduction, and cellular homing of immune responsive cells from the circulation to cites of increased antigen concentration.

In some instances, the diversified nature of immune cell function can be attributed to the specific pattern of expression of such cell surface proteins. For example, cells expressing VCAM proteins of the Ig superfamily are known to be involved primarily in cellular adhesion, whereas T lymphocytes characteristically express distinct patterns of the phenotypic markers, CD4, CD3, and CD8. Given importance of such cell surface proteins in the proper functioning of the immune system, there exists a need to identify novel cell-surface molecules which function to regulate the immune response and whose aberrant function can lead to immune response disorders such as congenital or acquired immunodeficiency, and or inflammatory disorders such as arthritis.

The placenta is the source of several peptide hormones that are homologous to hormones synthesized in other endocrine tissues. These placental hormones, which belong to the prolactin-growth hormone superfamily, are believed to play crucial roles in normal fetal development. Members of the prolactin-growth hormone superfamily include mouse placental lactogen I (mPL-I), mouse placental lactogen II (mPL-II), which bind to the prolactin receptor, and other proteins like mouse proliferin (PLF) (Linzer D.I.H. et al., *Proc. Natl. Acad. Sci. U.S.A.* (1985) 82:4356; Lee S.J. et al., *Endocrinology* (1988) 122:1761), mouse proliferin-related protein (PRP) (Linzer D.I.H. and Nathans D., *EMBO J.* 4:1419; Colosi P. et al., *Mol. Endocrinol.* (1988) 2:579), rat PL-I variant (Deb S. et al., *J. Biol. Chem.* (1991) 266:1605-1610), and rat PRL-like proteins (PLP) A (Campbell W.J. et al., *Endocrinology* (1989) 125:1565-1574), B (Ogilvie S. et al., *Endocrinology* (1990) 126:2561-2566), and C (Deb S. et al., *Endocrinology* (1991) 128:3066-3072; Deb S. et al., *J. Biol. Chem.* (1991) 266:23027-23032). PLF was discovered as a serum growth factor-inducible mRNA (Linzer D.I.H.

and Nathans D., *Proc. Natl. Acad. Sci. U.S.A.* (1983) 80:4271; *ibid.* (1984) 81:4255) and protein (Nilsen-Hamilton M. et al., *Cell* (1980) 20:19) in mouse fibroblasts, and expression of PLF in muscle cells has been shown to inhibit muscle cell-specific gene expression and differentiation (Wilder E.L. and Linzer D.I.H., *Mol. Cell. Biol.* (1989) 9:430; Muscat G.E.O. et al., *Mol. Endocrinol.* (1991) 5:802). The PRP mRNA was detected in placenta as a cDNA clone which cross-hybridized to the PLF cDNA (Linzer D.I.H. and Nathans D., *EMBO J.* 4:1419). All of these proteins show significant structural similarity (Southard J.N., *Molecular and Cellular Endocrinoilogy* (1991), 79: C133-C140) and are produced by the same trophoblast giant cells (Yamaguchi M., *Program of the 75th Annual Meeting of the Endocrine Society*, Las Vegas (1993), p.113 abstract), but their biological activities and gestational profiles in the maternal blood differ. The known biological activities of mPL-I and mPL-II are prolactin-like. The functions of PRP and PLF are not understood, but both have been postulated to be involved in regulating the initiation and then the cessation of placental neovascularization (Jackson D. et al., *Science* (1994) 266:1581-1584).

The differentiation of hematopoietic stem cells (HSCs) involves a series of lineage commitment steps accompanied by the acquisition of specific phenotypic characteristics (Huang, S. and Terstappen, L. (1992) *Nature* 360: 745-49). Cells gain or lose antigenic features and responsiveness to specific cytokines and growth factors based on their lineage and stage of differentiation. As development proceeds, HSCs become committed to specific myeloid, lymphoid or erythroid lineages. These committed "progenitor" stem cells ultimately differentiate into a wide variety of specialized cell types which include erythrocytes, neutrophils, basophils, eosinophils, platelets, mast cells, monocytes, tissue macrophages, osteoclasts, and the T and B lymphocytes.

Red blood cells (erythrocytes), white blood cells (leucocytes) and platelets (thrombocytes) are the predominant cell-types in the blood. Platelets are derived from detached fragments of larger cells called megakaryocytes which reside predominantly in the bone marrow. Megakaryopoiesis and platelet production are central to the release of cytokines, wound healing and blood coagulation. The failure of an organism to maintain adequate megakaryocyte numbers lends to thrombocytopenia and consequent bleeding disorders that can, in the extreme, result in death. Several humoral factors have been shown to promote megakaryocyte and platelet development, including interleukin-1 (IL-1) (Schmidt, J.A., *J. Exp. Med* 160:772-787, 1984; March, C.J. et al., *Nature* 315:641-647, 1985); IL-3 (Yang, Y.C. et al., *Cell* 47:3-10, 1986; Ikebuchi, K. et al., *Proc. Natl. Acad. Sci. USA* 84:9035-9039, 1987), IL-6 (Hirano, T., et al., *Proc. Natl. Acad. Sci. USA* 82:5490-5494, 1985; Hirano, T., et al., *Nature* 324:73-76,1986; Ishibashi, T. et al., *Proc. Natl. Acad. Sci. USA* 86:5953-5957, 1989); IL-11 (Paul, S.R. et al., *Proc. Natl. Acad.*

Sci. USA 87:7512-7516, 1990; Teramura, M. et al., *Blood* 79:327-331, 1992), leukemia inhibitory factor (Metcalf, D. et al., *Blood* 77:2150-2153, 1991), granulocyte-macrophage colony-stimulating factor (Wong, G., et al., *Science* 228:810-815, 1985), erythropoietin (Miyake, T. et al., *J. Biol. Chem.* 252:5558-5564, 1977; Jacobs, K. et al., *Nature* 313:806-815, 1985), and stem cell factor (Hendrie, P.C. et al., *Exp. Hematol.* 19:1031-1037, 1991). However, most of these factors are pleiotropic and consequently their roles in physiological replication of thrombocyte poiesis are unclear. Additional activities implicated in megakaryopoiesis include megakaryocyte-potentiating factor (Yamaguchi, N. et al., *J. Biol. Chem.* 269:805-808, 1994), megakaryocyte stimulatory factor (Tayrien, G., and Rosenberg, R.D., *J. Biol. Chem.* 262:3262-3268, 1987; Greenberg, S.M. et al., *Exp. Hematol.* 19:1031-1037, 1987), megakaryocyte colony-stimulating factor (Ogata, K. et al., *Int. J. Cell Cloning* 8:103-120, 1990; Erikson-Miller, C.L. et al., *Blood Cell Growth Factors: Their Present and Future Use in Hematology and Oncology*, Alpha Med Press, pp. 204-220, 1992; Erikson-Miller, C.L. et al., *Br. J. Haematol.* 84:197-203, 1993), thrombopoiesis-stimulating factor (McDonald, T.P. et al., *J. Lab. Clin. Med.* 85:59-66, 1975), and thrombopoietin (Hill, R. and Levin, *J., Exp. Hematol.* 14:752-759, 1986; Hill, R.J. et al., *J. Exp. Hematol* 20:354-360, 1992). Sources for these activities have included the urine, serum or plasma from aplastic and/or thrombocytopenic humans (McDonald, T.P., *Biochem. Med.* 13:101-110, 1975; Ogata, K. et al., *Int. J. Cell Cloning* 8:103-120, 1990), rats (Odell, T.T. et al., *Proc. Soc. Exp. Biol. Med.* 108:428-431, 1961), rabbits (Evatt, B.L. et al., *J. Lab. Clin. Med.* 83:364-371, 1974; Hill, R.J. et al., *Exp. Hematol.* 20:354-360, 1992), and dogs (Mazur, E. and South, K., *Exp. Hematol.* 13:1164-1172, 1985).

Recent studies have implicated the ligand of the *c-mpl* cytokine receptor, thrombopoietin (TPO), as a megakaryocyte lineage-specific factor (Bartley, T.D. *et al.* 1994; *Cell* Vol. 77:1117-1124; de Sauvage, F.J., *et al.* (1994) Vol. 369:533-38; Gurney, A.L. (1995) *Blood* Vol. 85 (4):981-88; Sohma, Y. et al. (1994) *FEBS Letters* 353:47-61). The human TPO cDNA encodes a mature protein of approximately 332 amino acids that can be divided into two domains: an amino terminal domain of 153 amino acids with homology to erythropoietin and a unique C-terminal domain of 175 amino acids containing multiple N-linked glycosylation sites. Both recombinant full-length human TPO and a truncated form consisting of the EPO-like domain (TPO_{153}) stimulate [^{3}H]-thymidine incorporation in murine BaF3 cells transfected with human *c-mpl,* demonstrating that he epo-like domain alone is sufficient for activation of *c-mpl* (deSauvage F.J. et al. (1994) *Nature* 369:533). Recombinant human TPO stimulated human megakaryocytopoiesis *in vitro* alone or in the presence of other exogenously added early-acting hematopoietic growth factors (deSauvage F.J. et al., *Nature* 369:533,

1994; Bartley, T.D. et al., *Cell* 77:1117, 1994; Kaushansky K. et al., *Nature* 369:568, 1994; Wendling F., *Nature* 369:571, 1994) Also, TPO stimulated platelet production in mice and dramatically increased the number of megakaryocytes in the spleen and bone marrow, indicating that TPO regulates thrombopoiesis *in vivo* (deSauvage F.J. et al., *Nature* 369:533, 1994; Lok S. et al., *Nature* 369:565, 1994; Kaushansky K. et al., *Nature* 369:568, 1994; Wendling F., *Nature* 369:571, 1994).

Summary of the Invention

The present invention is based, at least in part, on the discovery of novel signal peptide containing molecules referred to herein as Leukocyte-Specific Protein-1 ("LSP-1") , Proliferin Analog I ("PA-I"), and "Thrombopoietin Analog Protein" ("TAP-1") nucleic acid and protein molecules. The LSP-1, PA-I, and TAP-1 molecules of the present invention are useful as modulating agents in regulating a variety of cellular processes. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding LSP-1, PA-I, and TAP-1 proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of LSP-1, PA-I, and TAP-1-encoding nucleic acids. In one embodiment, the LSP-1, PA-I, and TAP-1 nucleic acid molecule is a naturally occurring nucleotide sequence.

In one embodiment, a LSP-1, PA-I, or TAP-1 nucleic acid molecule of the invention is at least 46%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or more identical to the nucleotide sequence (e.g., to the entire length of the nucleotide sequence) shown in SEQ ID NO:1, 3, 4, 6, 7, or 9 or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98554 or _____, or a complement thereof. In a preferred embodiment, a LSP-1 nucleic acid molecule of the invention is at least 46%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or more identical to the nucleotide sequence (e.g., to the entire length of the nucleotide sequence) shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a complement thereof.

In a preferred embodiment, the isolated nucleic acid molecule includes the nucleotide sequence shown SEQ ID NO:1 or 3, or a complement thereof. In another embodiment, the nucleic acid molecule includes SEQ ID NO:3 and nucleotides 1-1331 of SEQ ID NO:1. In another embodiment, the nucleic acid molecule includes SEQ ID NO:3 and nucleotides 2010-2462 of SEQ ID NO:1. In another preferred embodiment, the nucleic acid molecule consists of the nucleotide sequence shown in SEQ ID NO:1 or 3. In another preferred embodiment, the nucleic acid molecule includes a fragment of at

least 601 nucleotides of the nucleotide sequence of SEQ ID NO:1, 3, or a complement thereof.

In another preferred embodiment, the isolated nucleic acid molecule includes the nucleotide sequence shown SEQ ID NO:4 or 6, or a complement thereof. In another embodiment, the nucleic acid molecule includes SEQ ID NO:6 and nucleotides 1-54 of SEQ ID NO:4. In another embodiment, the nucleic acid molecule includes SEQ ID NO:6 and nucleotides 814-933 of SEQ ID NO:4. In another preferred embodiment, the nucleic acid molecule consists of the nucleotide sequence shown in SEQ ID NO:4 or 6.

In a preferred embodiment, the isolated nucleic acid molecule includes the nucleotide sequence shown SEQ ID NO:7 or 9, or a complement thereof. In another embodiment, the nucleic acid molecule includes SEQ ID NO:9 and nucleotides 259-523 of SEQ ID NO:7. In another preferred embodiment, the nucleic acid molecule consists of the nucleotide sequence shown in SEQ ID NO:7 or 9.

In another embodiment, a LSP-1, PA-I, and TAP-1 nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2, 5, or 8 or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98554 or _____. In a preferred embodiment, a LSP-1, PA-I, and TAP-1 nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 52%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more homologous to the entire length of the amino acid sequence of SEQ ID NO:2, 5, or 8 or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98554 or _____. In another preferred embodiment, a LSP-1 nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 52%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more homologous to the entire length of the amino acid sequence of SEQ ID NO:2 or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

In another preferred embodiment, an isolated nucleic acid molecule encodes the amino acid sequence of human LSP-1, PA-I, and TAP-1. In yet another preferred embodiment, the nucleic acid molecule includes a nucleotide sequence encoding a protein having the amino acid sequence of SEQ ID NO: 2, 5, or 8 or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98554 or _____. In yet another preferred embodiment, the nucleic acid molecule is at least 601 nucleotides in length. In a further preferred embodiment, the nucleic acid molecule is at least 601 nucleotides in length and encodes a protein having a LSP-1, PA-I, and TAP-1 activity (as described herein).

Another embodiment of the invention features nucleic acid molecules, preferably LSP-1, PA-I, and TAP-1 nucleic acid molecules, which specifically detect LSP-1, PA-I, and TAP-1 nucleic acid molecules relative to nucleic acid molecules encoding non-LSP-1, non-PA-I, and non-TAP-1 proteins. For example, in one embodiment, such a nucleic acid molecule is at least 300-350, 350-400, 400-450, 450-500, 500-550, 550-600, 601 or more nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:1, 4, or 7, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98554 or _____, or a complement thereof.

In other preferred embodiments, the nucleic acid molecule encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 5, or 8 or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98554 or _____, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, 3, 4, 6, 7, or 9 under stringent conditions.

Another embodiment of the invention provides an isolated nucleic acid molecule which is antisense to a LSP-1, PA-I, and TAP-1 nucleic acid molecule, e.g., the coding strand of a LSP-1, PA-I, and TAP-1 nucleic acid molecule.

Another aspect of the invention provides a vector comprising a LSP-1, PA-I, and TAP-1 nucleic acid molecule. In certain embodiments, the vector is a recombinant expression vector. In another embodiment, the invention provides a host cell containing a vector of the invention. In yet another embodiment, the invention provides a host cell containing a nucleic acid molecule of the invention. The invention also provides a method for producing a protein, preferably a LSP-1, PA-I, and TAP-1 protein, by culturing in a suitable medium, a host cell, e.g., a mammalian host cell such as a non-human mammalian cell, of the invention containing a recombinant expression vector, such that the protein is produced.

In another embodiment, the invention features fragments of the protein having the amino acid sequence of SEQ ID NO: 2, 5, or 8, wherein the fragment comprises at least 15 amino acids (e.g., contiguous amino acids) of the amino acid sequence of SEQ ID NO: 2, 5, or 8 or an amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number 98554 or _____. In another embodiment, the protein, preferably a LSP-1, PA-I, or TAP-1 protein, has the amino acid sequence of SEQ ID NO: 2, 5, or 8, respectively.

In another embodiment, the invention features an isolated protein, preferably a LSP-1, PA-I, and TAP-1 protein, which is encoded by a nucleic acid molecule consisting of a nucleotide sequence at least about 46%, 50%, 55%, 60%, 65%, 70%,

75%, 80%, 85%, 90%, 95%, 98% or more homologous to a nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or a complement thereof. This invention further features an isolated protein, preferably a LSP-1, PA-I, or TAP-1 protein, which is encoded by a nucleic acid molecule consisting of a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or a complement thereof.

The proteins of the present invention or portions thereof, e.g., biologically active portions thereof, can be operatively linked to a non-LSP-1, non-PA-I, or non-TAP-1 polypeptide (e.g., heterologous amino acid sequences) to form fusion proteins. The invention further features antibodies, such as monoclonal or polyclonal antibodies, that specifically bind proteins of the invention, preferably LSP-1, PA-I, and TAP-1 proteins. In addition, the LSP-1, PA-I, and TAP-1 proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides a method for detecting the presence of a LSP-1, PA-I, and TAP-1 nucleic acid molecule, protein or polypeptide in a biological sample by contacting the biological sample with an agent capable of detecting a LSP-1, PA-I, and TAP-1 nucleic acid molecule, protein or polypeptide such that the presence of a LSP-1, PA-I, and TAP-1 nucleic acid molecule, protein or polypeptide is detected in the biological sample.

In another aspect, the present invention provides a method for detecting the presence of LSP-1, PA-I, and TAP-1 activity in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of LSP-1, PA-I, and TAP-1 activity such that the presence of LSP-1, PA-I, and TAP-1 activity is detected in the biological sample.

In another aspect, the invention provides a method for modulating LSP-1, PA-I, and TAP-1 activity comprising contacting a cell capable of expressing LSP-1, PA-I, and TAP-1 with an agent that modulates LSP-1, PA-I, and TAP-1 activity such that LSP-1, PA-I, and TAP-1 activity in the cell is modulated. In one embodiment, the agent inhibits LSP-1, PA-I, and TAP-1 activity. In another embodiment, the agent stimulates LSP-1, PA-I, and TAP-1 activity. In one embodiment, the agent is an antibody that specifically binds to a LSP-1, PA-I, and TAP-1 protein. In another embodiment, the agent modulates expression of LSP-1, PA-I, and TAP-1 by modulating transcription of a LSP-1, PA-I, and TAP-1 gene or translation of a LSP-1, PA-I, and TAP-1 mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of a LSP-1, PA-I, and TAP-1 mRNA or a LSP-1, PA-I, and TAP-1 gene.

In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by aberrant LSP-1, PA-I, and TAP-1 protein or nucleic acid expression or activity by administering an agent which is a LSP-1, PA-I, and TAP-1 modulator to the subject. In one embodiment, the LSP-1, PA-I, and TAP-1 modulator is a LSP-1, PA-I, and TAP-1 protein. In another embodiment the LSP-1, PA-I, and TAP-1 modulator is a LSP-1, PA-I, and TAP-1 nucleic acid molecule. In yet another embodiment, the LSP-1, PA-I, and TAP-1 modulator is a peptide, peptidomimetic, or other small molecule. In a preferred embodiment, the disorder characterized by aberrant LSP-1, PA-I, and TAP-1 protein or nucleic acid expression is a disorder characterized by deregulated angiogenesis, deregulated immune response, or deregulated hematopoiesis.

The present invention also provides a diagnostic assay for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding a LSP-1, PA-I, and TAP-1 protein; (ii) mis-regulation of the gene; and (iii) aberrant post-translational modification of a LSP-1, PA-I, and TAP-1 protein, wherein a wild-type form of the gene encodes a protein with a LSP-1, PA-I, and TAP-1 activity.

In another aspect the invention provides a method for identifying a compound that binds to or modulates the activity of a LSP-1, PA-I, and TAP-1 protein, by providing an indicator composition comprising a LSP-1, PA-I, and TAP-1 protein having LSP-1, PA-I, and TAP-1 activity, contacting the indicator composition with a test compound, and determining the effect of the test compound on LSP-1, PA-I, and TAP-1 activity in the indicator composition to identify a compound that modulates the activity of a LSP-1, PA-I, and TAP-1 protein.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 depicts the cDNA sequence and predicted amino acid sequence of human LSP-1. The nucleotide sequence corresponds to nucleic acids 1 to 2462 of SEQ ID NO:1. The amino acid sequence corresponds to amino acids 1 to 226 of SEQ ID NO:2, 5, or 8.

Figure 2 is a schematic drawing depicting selected clones which were isolated and sequenced to derive the nucleotide sequence of the gene encoding human LSP-1. The figure details the relationship between the original, partial LSP-1 clone isolated from a bone marrow cDNA library, three additional clones for which partial sequence information was available, and the final composite sequence generated from complete

sequence analysis of the additional clones as well as the nucleotide sequence of the original partial clone.

Figure 3 is a schematic diagram depicting the biological and functional domains of human LSP-1. The LSP-1 protein comprises at least a signal peptide from about amino acids 1-20 of the amino acid sequence depicted (which corresponds to the amino acid sequence of SEQ ID NO:2, 5, or 8), an Ig-like domain from about amino acids 46-128 of the amino acid sequence depicted, and a transmembrane domain from about amino acids 192-213 of the amino acid sequence depicted.

Figure 4 depicts the cDNA sequence and predicted amino acid sequence of murine Proliferin analog I. The nucleotide sequence corresponds to nucleic acids 1 to 933 of SEQ ID NO:4. The amino acid sequence corresponds to amino acids 1 to 253 of SEQ ID NO:5.

Figure 5 depicts an alignment of the amino acid sequences of murine Proliferin analog I (corresponding to amino acids 1 to 247 of SEQ ID NO:5), and murine proliferin related protein (Swiss-Prot™ Accession No. P04769).

Figure 6 depicts northern blots performed using clone aa014234 as a probe. Figure 6A depicts a northern blot using human tissue from placenta, heart, brain, lung, liver, skeletal muscle, kidney and pancreas. Figure 6B depicts a northern blot using tissue from mouse embryos (day 7, 11, 15, and 17 embryos). Figure 6C depicts a northern blot using human tissue from brain, lung, liver, and kidney.

Figure 7 depicts the cDNA sequence and predicted amino acid sequence of human TAP-1. The nucleotide sequence corresponds to nucleic acids 1 to 528 of SEQ ID NO:7. The amino acid sequence corresponds to amino acids 1 to 86 of SEQ ID NO:8.

Figure 8 depicts an alignment of the amino acid sequences of human TAP-1 (corresponding to amino acids 15 to 75 of SEQ ID NO:8) and human TPO (Swiss-Prot™ Accession Numbers P40225, 1401246, 939627). Identical amino acids are indicated by a single amino acid code in the row between the TAP-1 and TPO sequences, conserved amino acids are indicated as (+).

Figure 9 depicts an alignment of the LSP-1 nucleic acid molecule with theFDF03 nucleic acid molecule (described in WO/24906) using the GAP program in the GCG software package (pam120 matrix) and a gap weight of 12 and a length weight of 4.

Figure 10 depicts an alignment of the LSP-1 protein with theFDF03 protein (described in WO/24906) using the GAP program in the GCG software package (pam120 matrix) and a gap weight of 12 and a length weight of 4.

Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of novel molecules specific to peripheral blood leukocytes, referred to herein as LSP-1 protein and nucleic acid molecules, which comprise a family of molecules having certain conserved structural and functional features. The present invention is further based, at least in part, on the discovery of novel molecules of the prolactin-growth hormone superfamily, referred to herein as "Proliferin analog I" or "PA-I" protein and nucleic acid molecules, which comprise a family of molecules having certain conserved structural and functional features. Moreover, the invention is based, at least in part, on the discovery of novel hematopoietic specific factors, referred to herein as "Thrombopoietin Analog Protein" or "TAP-1" nucleic acid and protein molecules.

The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more proteins or nucleic acid molecules having a common structural domain and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin and a homologue of that protein of murine origin, as well as a second, distinct protein of human origin and a murine homologue of that protein. Members of a family may also have common functional characteristics.

One embodiment of the invention features LSP-1, PA-I, and TAP-1 nucleic acid and protein molecules, preferably human LSP-1, PA-I, and TAP-1 nucleic acid and protein molecules. The LSP-1, PA-I, and TAP-1 nucleic acid and protein molecules of the invention are described in further detail in the following subsections.

A. The LSP-1 Nucleic Acid and Protein Molecules

In one embodiment, a LSP-1 family member is identified based on the presence of "an immunoglobulin (Ig)-like domain" and a "transmembrane domain" in the protein or corresponding nucleic acid molecule. As used herein, the term "immunoglobulin (Ig)-like domain" refers to a protein domain having an amino acid sequence of at least about 50, preferably at least about 60, more preferably at least about 70 amino acid residues, and even more preferably at least about 80-90 amino acids of which at least about 30%, preferably at least about 40%, more preferably at least about 50%, 60% or 70% of the amino acids are homologous to the amino acid sequence of an immunoglobulin domain. The homologous amino acids between an Ig-like domain and an Ig domain can be positioned across the entire Ig-like domain, referred to herein as a "full" Ig-like domain. Alternatively, the homologous amino acids between an Ig-like domain and an Ig domain can be concentrated in regions of high homology dispersed

throughout the Ig-like domain, referred to as a "partial" Ig-like domain. In a preferred embodiment, an Ig-like domain is located in the N-terminal region of a LSP-1 protein. For example, in one embodiment, a LSP-1 protein contains an Ig-like domain containing about amino acids 46-128 of SEQ ID NO:2, wherein amino acids homologous to an Ig domain are concentrated between amino acids 46-78 and amino acids 109-128. In another embodiment an LSP-1 protein includes an Ig-like domain or a partial Ig-like domain which is at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or more homologous to the Ig-like domain of SEQ ID NO:2.

Also as used herein, a "transmembrane domain" refers to a protein domain having an amino acid sequence containing at least about 10, preferably about 13, preferably about 16, more preferably about 19, and even more preferably about 21, 23, 25, 30, 35 or 40 amino acid residues, of which at least about 60-70%, preferably about 80% and more preferably about 90% of the amino acid residues contain non-polar side chains, for example, alanine, valine, leucine, isoleucine, proline, phenylalanine, tryptophan, and methionine. A transmembrane domain is lipophillic in nature. In a preferred embodiment, a transmembrane domain is located in the C-terminal region of a LSP-1 protein. For example, in one embodiment, a LSP-1 protein contains a transmembrane domain containing about amino acids 192-213 of SEQ ID NO:2.

In another preferred embodiment, a LSP-1 family member is identified further based on the presence of an "N-terminal signal sequence". As used herein, a "signal sequence" refers to a peptide containing about 20 amino acids which occurs at the extreme N-terminal end of secretory and integral membrane proteins and which contains large numbers of hydrophobic amino acid residues. Such a "signal sequence", also referred to in the art as a "signal peptide", serves to direct a protein containing such a sequence to a lipid bilayer. For example, in one embodiment, an LSP-1 protein contains a signal sequence containing about amino acids 1-20 of SEQ ID NO:2.

Preferred LSP-1 molecules of the present have an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2. As used herein, the term "sufficiently homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains have at least about 40% homology, preferably 50% homology, more preferably 60%-70% homology across the amino acid sequences of the domains and

contain at least one, preferably two, more preferably three, and even more preferably four, five or six structural domains, are defined herein as sufficiently homologous. Furthermore, amino acid or nucleotide sequences which share at least 40%, preferably 50%, more preferably 60, 70, or 80% homology and share a common functional activity are defined herein as sufficiently homologous.

As used interchangeably herein a "LSP-1 activity", "biological activity of LSP-1" or "functional activity of LSP-1", refers to an activity exerted by a LSP-1 protein, polypeptide or nucleic acid molecule on a LSP-1 responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. In one embodiment, a LSP-1 activity is a direct activity, such as an association with or an enzymatic activity on a second protein. In another embodiment, a LSP-1 activity is an indirect activity, such as a cellular signaling activity mediated by interaction of the LSP-1 protein with a second protein. In a preferred embodiment, a LSP-1 activity is at least one or more of the following activities: (i) interaction of a LSP-1 protein on the cell surface with a second non-LSP-1 protein molecule on the surface of the same cell; (ii) interaction of a LSP-1 protein on the cell surface with a second non-LSP-1 protein molecule on the surface of a different cell; (iii) complex formation between a soluble LSP-1 protein and a cognate ligand; (iv) complex formation between a membrane-bound LSP-1 protein and a cytokine; (v) interaction of a LSP-1 protein with an intracellular protein via a second protein on the cell surface. In yet another preferred embodiment, a LSP-1 activity is at least one or more of the following activities: (i) modulation of cellular signal transduction; (ii) regulation of a cell involved in an inflammatory response; (iii) homing of a cell having a LSP-1 protein on its cell surface from a first to a second anatomical location; and (iv) modulation of a cell involved in the immune response.

Accordingly, another embodiment of the invention features isolated LSP-1 proteins and polypeptides having a LSP-1 activity. Preferred LSP-1 proteins have at least an N-terminal partial immunoglobulin (Ig) domain, a C-terminal transmembrane domain, and a LSP-1 activity. In another preferred embodiment, the LSP-1 protein has at least at least an N-terminal partial immunoglobulin (Ig) domain, a C-terminal transmembrane domain, a LSP-1 activity, and an amino acid sequence sufficiently homologous to an amino acid sequence of SEQ ID NO:2.

In a particularly preferred embodiment, the LSP-1 protein and nucleic acid molecules of the present invention are human LSP-1 molecules. A human LSP-1 cDNA was identified by the Signal Peptide trAP methodology described herein. (See Example 1). The nucleic acid sequence of a positive cDNA clone was used to search the GenBank™ EST database and multiple ESTs having greater than 95% nucleotide identity were found. Three clones containing published nucleotide sequences were

purchased from Research Genetics (Huntsville, AL) as part of the IMAGE Consortium. The full sequence of the human clone was assembled as depicted in Figure 2. A nucleotide sequence of the isolated human LSP-1 cDNA and the predicted amino acid sequence of the human LSP-1 protein are shown in Figure 1 and in SEQ ID NOs:1 and 2, respectively. In addition, the nucleotide sequence corresponding to the coding region of the human LSP-1 cDNA (nucleotides 1332-2009) is represented as SEQ ID NO:3.

An approximately 1.5 kb LSP-1 mRNA transcript is expressed at very low levels in most human tissues tested. Significant expression of LSP-1 mRNA was detected only in peripheral blood leukocytes. Chromosomal mapping indicates that the human LSP-1 gene maps to chromosome 7q21-q22, at 111-112 cM.

The human LSP-1 cDNA, which is approximately 2462 nucleotides in length, encodes a protein which is approximately 226 amino acid residues in length. The human LSP-1 protein contains an N-terminal signal sequence, an N-terminal partial immunoglobulin (Ig) domain, and a C-terminal transmembrane domain. A LSP-1 N-terminal partial immunoglobulin (Ig) domain can be found at least, for example, from about amino acids 46-128 of SEQ ID NO:2. A LSP-1 C-terminal transmembrane can be found at least, for example, from about amino acids 192-213 of SEQ ID NO:2. The human LSP-1 protein is a membrane-bound protein which further contains a signal sequence at about amino acids 1-20 of SEQ ID NO:2. The prediction of such a signal peptide can be made, for example, utilizing the computer algorithm SIGNALP (Henrik, et al. (1997) *Protein Engineering* 10:1-6).

B. <u>The PA-I Nucleic Acid and Protein Molecules</u>

In another embodiment, a PA-I family is identified based on the presence of at least one "cysteine-rich domain" in the protein or corresponding nucleic acid molecule. As used herein, the term "cysteine rich domain" includes a protein domain having an amino acid sequence of at least about 20 amino acids of which at least about 2 amino acids are cysteine residues. Preferably, a cysteine rich domain includes at least about 30, more preferably at least about 35-40 amino acid residues, of which at least about 2, preferably at least about 3, more preferably at least about 4, 5 or 6 amino acids are cysteine residues. Cysteine-rich domains having lengths of 45-50 or 60 amino acid residues and having up to 7, 8, 9 or 10 cysteine residues are also within the scope of this invention. Cysteine rich domains are described in, for example, Lodish H. et al. Molecular Cell Biology, (Scientific American Books Inc., New York, N.Y., 1995), the contents of which are incorporated herein by reference. In one embodiment, a PA-I protein includes a cysteine rich domain having at least about 20%, preferably at least about 30%, and more preferably about 40% amino acid sequence homology to a

proliferin-related protein cysteine-rich domain, such as the cysteine-rich domain of SEQ ID NO:10 (e.g., amino acid residues 198-243 of murine proliferin-related protein, Swiss-Prot™ Accession No. P04769).

Preferred PA-I molecules of the invention have an amino acid sequence sufficiently homologous to an amino acid sequence of SEQ ID NO:5. As used herein, the term "sufficiently homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have a common structural domain and/or common functional activity. For example, amino acid or nucleotide sequences which contain a common structural domain having about 40% homology, preferably 50% homology, more preferably 60%-70% homology are defined herein as sufficiently homologous. In one embodiment, the PA-I protein contains a cysteine-rich domain and a PA-I activity.

As used interchangeably herein a "PA-I activity", "biological activity of PA-I" or "functional activity of PA-I", refers to an activity exerted by a PA-I protein, polypeptide or nucleic acid molecule on a PA-I responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. In one embodiment, a PA-I activity is a direct activity, such as an association with or an enzymatic activity on a second protein. In another embodiment, a PA-I activity is an indirect activity, such as a cellular signaling activity mediated by interaction of the PA-I protein with a second protein. In a preferred embodiment, a PA-I activity is at least one or more of the following activities: (i) formation of a complex with a cell-surface protein(s) or a ligand, e.g., a lipid or carbohydrate; (ii) formation of a complex with a prolactin and/or growth hormone receptor. In yet another preferred embodiment, a PA-I activity is at least one or more of the following activities: (i) regulation of cellular growth; (ii) regulation of cellular proliferation; (iii) regulation of angiogenesis; (iv) regulation of cellular differentiation; and (v) regulation of cell survival.

Accordingly, another embodiment of the invention features isolated PA-I proteins and polypeptides having a PA-I activity. Preferred PA-I proteins have at least one cysteine-rich domain and a PA-I activity. In another preferred embodiment, the PA-I protein has at least one cysteine-rich domain, a PA-I activity and an amino acid sequence sufficiently homologous to an amino acid sequence of SEQ ID NO:5.

Yet another embodiment of the invention features PA-I molecules which contain a signal sequence. As used herein, a "signal sequence" refers to a peptide containing about 20 amino acids which occurs at the extreme N-terminal end of secretory and integral membrane proteins and which contains large numbers of hydrophobic amino

acid residues. Such a "signal sequence", also referred to in the art as a "signal peptide", serves to direct a protein containing such a sequence to a lipid bilayer.

In a further embodiment, the invention features PA-I molecules which are secreted. As used herein, "secreted" refers to protein molecules which have the ability to be directed to the cellular plasma membrane (usually through a signal peptide) and subsequently released into the extracellular space. Such secreted PA-I molecules lack a trasmembrane domain.

In a particularly preferred embodiment, the PA-I protein and nucleic acid molecules of the present invention are human PA-I molecules. Using clone aa014234 as a probe for northern blots, a band of 1Kb was detected in human placenta tissue (see Figure 6).

In another preferred embodiment, the PA-I protein and nucleic acid molecules of the present invention are murine PA-I molecules. A murine PA-I cDNA (also refered to as MOPAI or TANGO 95) was identified by homology with human growth hormone. In particular, a human growth hormone cDNA sequence was used to search the GenBank™ EST database and clone aa014234 was identified. This clone, containing the published nucleotide sequence, was purchased from Research Genetics (Huntsville, AL) as part of the IMAGE Consortium and subsequently fully sequenced. A nucleotide sequence of the isolated murine PA-I cDNA and the predicted amino acid sequence of the murine PA-I protein are shown in Figure 4 and in SEQ ID NOs:4 and 5, respectively. In addition, the nucleotide sequence corresponding to the coding region of the murine PA-I cDNA (nucleotides 55-816) is represented as SEQ ID NO:6.

A BlastP search (BLAST™ searching, utilizing an amino acid sequence against a protein database), using the translation product (frame 1) of the cDNA sequence represented as SEQ ID NO:4, revealed homology to proteins belonging to the prolactin-growth hormone superfamily. One example of such a protein is mouse proliferin-related protein, which is 243 amino acids in length and is 35% identical (see Figure 5) to amino acids 1-247 of the murine PA-I amino acid sequence depicted in Figure 4 and SEQ ID NO:5.

A 1 kb PA-I mRNA transcript is expressed in murine tissues from day 7 embryos (see Figure 6).

The murine PA-I cDNA, which is 933 nucleotides in length, encodes a protein which is approximately 253 amino acid residues in length. The murine PA-I protein contains one cysteine-rich domain. A PA-I cysteine rich domain can be found at least, for example, from about amino acids 201-247 of SEQ ID NO:5 (Lys201 to Lys247 of the murine PA-I amino acid sequence). The murine PA-I protein is a secreted protein which lacks a transmembrane domain. The murine PA-I protein further contains a

signal sequence at amino acids 1-30 of SEQ ID NO:5 (Met1 to Ser30 of the murine PA-I amino acid sequence. The prediction of such a signal peptide can be made utilizing the computer algorithm SIGNALP (Henrik, et al. (1997) *Protein Engineering* 10:1-6).

C. **The TAP-1 Nucleic Acid and Protein Molecules**

The carboxy-terminal domain of TAP-1 molecules has homology with the carboxy-terminal domain of human thrombopoietin (TPO). TPO has been identified as the ligand of the *c-mpl* cytokine receptor which upon activation of the receptor functions as a megakaryocyte lineage-specific factor. The N-terminal domain of human TPO shares homology to erythropoietin (EPO). Thus, TAP-1, TPO and EPO may comprise a family of structurally- and functionally related factors. TAP-1 molecules of the present invention may influence cell proliferation or differentiation, for example, hematopoietic cell proliferation or differentiation, e.g., the maturation or differentiation of, megakaryocytes into platelets, or erythroid progenitor cells into erythrocytes.

In one embodiment, a TAP-1 family is identified based on the presence of at least one "serine-proline-threonine rich" in the protein or corresponding nucleic acid molecule. As used herein, the term "serine-proline-threonine rich" refers to a protein domain of about 7, 10, 15, 20, 30, 40, 50, 60, 70, or 80 amino acids, preferably 10 to 30 amino acids, and most preferably 18-22 amino acids having at least about 15% serine, proline and/or threonine residues, more preferably about 20 amino acids having at least about 20% serine, proline and/or threonine residues, and even more preferably about 20 amino acids having at least about 30% serine, proline and/or threonine residues.

In one embodiment, a TAP-1 protein includes a serine-proline-threonine rich domain having at least about 20%, preferably at least about 30%, and more preferably about 40% amino acid sequence homology to a TAP-1 serine-proline-threonine rich domain, such as the domain of SEQ ID NO:8 (e.g., amino acid residues 1-20 or 40-60 of SEQ ID NO:8).

Preferred TAP-1 molecules of the present have an amino acid sequence sufficiently homologous to all or a portion of the amino acid sequence of SEQ ID NO:8, such as a serine-proline-threonine rich domain of the amino acid sequence of SEQ ID NO:8. As used herein, the term "sufficiently homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have a common structural domain and/or common functional activity. For example, amino acid or nucleotide sequences which contain a common structural domain having about 40% homology, preferably 50%

homology, more preferably 60%-70% homology are defined herein as sufficiently homologous. In one embodiment, the a TAP-1 protein contains a serine-proline-threonine rich domain and a TAP-1 activity.

In one embodiment, a TAP-1 family is identified by protein which include a unique carboxyl terminal domain. The terms "C-terminal unique domain" or "carboxy-terminal domain" as used herein, refer to a protein domain of a TAP-1 protein family member which includes at least one serine-proline-threonine-rich domain and shares structural similarity to a human TPO C-terminal domain. A C-terminal unique domain is sufficiently homologous between TAP-1 protein family members such that the domain is at least about 40%, preferably about 50%, more preferably about 60%, even more preferably about 70%, 80%, or 90% homologous. As defined herein, a C-terminal unique domain of a TAP-1 protein family member, however, is not sufficiently homologous to a C-terminal unique domain of a member of another protein family, such as a TPO protein family.

As used interchangeably herein a "TAP-1 activity", "biological activity of TAP-1" or "functional activity of TAP-1", refers to an activity exerted by a TAP-1 protein, polypeptide or nucleic acid molecule on a TAP-1 responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. In one embodiment, a TAP-1 activity is a direct activity, such as an association with, or an enzymatic activity, on a second protein, e.g., a cell-surface receptor. In another embodiment, a TAP-1 activity is an indirect activity, such as a cellular signaling activity mediated by interaction of the TAP-1 protein with a second protein. In a preferred embodiment, a TAP-1 activity is at least one or more of the following activities: (i) interaction, e.g., binding to, a cell-surface receptor, e.g., a hematopoietic-cell surface receptor; (ii) modulation of, e.g., activation or inhibition of, a cell-surface receptor; (iii) modulation of cellular signal transduction. In yet another preferred embodiment, a TAP-1 activity is at least one or more of the following activities: (i) regulation of cellular proliferation; (ii) regulation of cellular differentiation; (iii) regulation of cell survival; (iv) modulation of a cell involved in the immune response (v) regulation of maturation and/or differentiation of a hematopoietic stem cell; (vi) modulation of megakaryocytopoiesis; (vii) modulation of thrombopoiesis; (viii) regulation of maturation and/or differentiation of a megakaryocyte into platelets; and (ix) regulation of maturation and/or differentiation of erythroid progenitor cells into erythrocytes.

Accordingly, another embodiment of the invention features isolated TAP-1 proteins and polypeptides having a TAP-1 activity. Preferred TAP-1 proteins have at least one serine-proline-threonine rich domain and a TAP-1 activity. In another preferred embodiment, the TAP-1 protein has at least one serine-proline-threonine rich

domain, a TAP-1 activity and an amino acid sequence sufficiently homologous to an amino acid sequence of SEQ ID NO:8.

In a particularly preferred embodiment, the TAP-1 protein and nucleic acid molecules of the present invention are human TAP-1 molecules. A partial human TAP-1 cDNA, also referred to as TANGO-94, was identified by analysis of an EST database using mouse TPO sequence as a probe. A partial human clone (jthqb070d08) was obtained from a human prostate cDNA library and was subsequently fully sequenced. Clone jthqb070d08 was deposited with the American Type Culture Collection on October 2, 1997 and has ATCC Accession Number 98554. This clone contains a nucleotide sequence of the isolated C-terminal domain of human TAP-1 cDNA (nucleotides 1-528 corresponding to the C-terminus and 3' untranslated sequence) and the predicted amino acid sequence of the human TAP-1 protein (amino acids 1-86) are shown in Figure 7 and in SEQ ID NOs:7 and 8, respectively. The amino acid sequences showed 32% identitity to the C-terminal part of human TPO. The nucleotide sequence corresponding to the coding region of the human TAP-1 cDNA are nucleotides 1-258 of SEQ ID NO:7, nucleotides 259-528 correspond to the 3' untranslated region of the gene.

Using jthqb070d08 cDNA as a probe, a 3 kb TAP-1 mRNA transcript is expressed in human fetal liver tissues. In addition to the 3 kb band, several bands were detected in the Northern blots which may indicate splice variants of TAP-1. Two other less intense bands of approximately 5 and 2 kb were detected in all tissues tested.

The partial human TAP-1 cDNA, which is approximately 528 nucleotides in length, and which is approximately 86 amino acid residues in length. The human TAP-1 protein contains four serine-proline-threonine-rich domains. A TAP-1 serine-proline-threonine-rich domain can be found at least, for example, from about amino acids 1-20 of SEQ ID NO:8 (Gly1 to Gly20 of SEQ ID NO:8); from about amino acids 21-40 of SEQ ID NO:8 (Ile20 to Ala40 of SEQ ID NO:8); from about amino acids 41-60 of SEQ ID NO:8 (Val40 to Gly60 of SEQ ID NO:8); and from about amino acids 61-81 of SEQ ID NO:8 (Pro61 to Thr81 of SEQ ID NO:8). The human TAP-1 C-terminal domain appears to encode a secreted protein, e.g., growth factor a secreted protein.

The human TAP-1 amino acid sequence shares significant homology, about 32% homology, with the C-terminal region of TPO. An alignment of the human TAP-1 amino acid sequences to human TPO sequences is presented in Figure 8. The figure depicts an alignment of the amino acid sequences of TAP-1 (corresponding to amino acids 15 to 75 of SEQ ID NO:8) and human TPO sequences (Swiss-Prot™ Accession Numbers P40225, 1401246, 939627). Identical residues are indicated in the row between the TAP-1 and the TPO sequences by a single amino acid code; conserved amino acid residues are indicated as (+).

Using jthqb070d08 cDNA as a probe, 8 clones from a human fetal liver library were isolated and submitted for sequencing. 3 out of the 8 clones contain an insert of approximately 3 kb.

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode LSP-1, PA-I, and TAP-1 proteins or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify LSP-1, PA-I, and TAP-1-encoding nucleic acids (e.g., LSP-1, PA-I, and TAP-1 mRNA) and fragments for use as PCR primers for the amplification or mutation of LSP-1, PA-I, and TAP-1 nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term "isolated nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated LSP-1, PA-I, and TAP-1 nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or a portion of these nucleotide molecules, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or portion of the nucleic acid sequences of SEQ ID NO:1, 3, 4, 6, 7, or 9 as a hybridization probe, LSP-1, PA-I, and

TAP-1 nucleic acid molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), .

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:1, 3, 4, 6, 7, or 9 can be isolated by the polymerase chain reaction using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9.

A nucleic acid molecule of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to LSP-1, PA-I, and TAP-1 nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:1. The sequence of SEQ ID NO:1 corresponds to the human LSP-1 cDNA. This cDNA comprises sequences encoding the human LSP-1 protein (i.e., "the coding region", from nucleotides 1332-2009), as well as 5' untranslated sequences (nucleotides 1 to 1331) and 3' untranslated sequences (nucleotides 2010 to 2462). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:1 (e.g., nucleotides 1332 to 2009, corresponding to SEQ ID NO:3).

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:4. The sequence of SEQ ID NO:4 corresponds to the murine PA-I cDNA. This cDNA comprises sequences encoding the murine PA-I protein (i.e., "the coding region", from nucleotides 55 to 816 of SEQ ID NO:4), as well as 5' untranslated sequences (nucleotides 1 to 54 of SEQ ID NO:4) and 3' untranslated sequences (nucleotides 817 to 933 of SEQ ID NO:4). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:4 (e.g., nucleotides 55 to 816 of SEQ ID NO:4, or SEQ ID NO:6).

In yet another preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:7. The sequence of SEQ ID NO:7 corresponds to the human TAP-1 cDNA. This cDNA comprises sequences encoding the human TAP-1 protein (i.e., "the coding region", from nucleotides 1 to 258 of SEQ ID NO:7, or SEQ ID NO:9), as well as 3' untranslated sequences (nucleotides 259 to 528 of SEQ ID NO:7). Alternatively, the nucleic acid

molecule can comprise only the coding region of SEQ ID NO:7 (e.g., nucleotides 1 to 258 of SEQ ID NO:7, or SEQ ID NO:9).

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or 9, or a portion of either of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or 9, is one which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or 9, such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or 9, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 60-65%, preferably at least about 70-75%, more preferable at least about 80-85%, and even more preferably at least about 90-95% or more homologous to the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or 9, or a portion of any of these nucleotide sequences.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of LSP-1, PA-I, and TAP-1. The nucleotide sequence determined from the cloning of the LSP-1, PA-I, and TAP-1 gene allows for the generation of probes and primers designed for use in identifying and/or cloning LSP-1, PA-I, and TAP-1 homologues in other cell types, e.g. from other tissues, as well as LSP-1, PA-I, and TAP-1 homologues from other mammals. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence which hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of SEQ ID NO:1, 3, 4, 6, 7, or 9, of an anti-sense sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or of a naturally occurring mutant of either SEQ ID NO:1, 3, 4, 6, 7, or 9.

Probes based on the LSP-1, PA-I, and TAP-1 nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a LSP-1, PA-I, and TAP-1 protein, such as by measuring a level of a LSP-1, PA-I, and TAP-1-encoding nucleic acid in a sample of cells from a subject e.g., detecting LSP-1, PA-I, and TAP-1 mRNA levels or determining whether a genomic LSP-1, PA-I, and TAP-1 gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of LSP-1, PA-I, and TAP-1" can be prepared by isolating a portion of SEQ ID NO:1, 3, 4, 6, 7, or 9, which encodes a polypeptide having a LSP-1, PA-I, and TAP-1 biological activity (the biological activities of the LSP-1, PA-I, and TAP-1 proteins are described herein), expressing the encoded portion of LSP-1, PA-I, and TAP-1 protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of LSP-1, PA-I, and TAP-1. For example, a nucleic acid fragment encoding a biologically active portion of PA-I encompasses at least nucleic acids 654-795 of SEQ ID NO:4 (encoding a murine PA-I cysteine rich domain).

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or 9 (and portions thereof) due to degeneracy of the genetic code and, thus, encode the same LSP-1, PA-I, and TAP-1 protein as that encoded by the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or 9. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:2, 5, or 8.

In addition to the LSP-1, PA-I, and TAP-1 nucleotide sequences shown in SEQ ID NO:1, 3, 4, 6, 7, and 9, respectively, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of LSP-1, PA-I, and TAP-1 may exist within a population (e.g., the human population). Such genetic polymorphism in the LSP-1, PA-I, and TAP-1 gene may exist among individuals within a population due to natural allelic variation.

As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a LSP-1, PA-I, and TAP-1 protein, preferably a mammalian LSP-1, PA-I, and TAP-1 protein, and can further include non-coding regulatory sequences, and introns.

Allelic variants of LSP-1, PA-I, and TAP-1 include both functional and non-functional LSP-1, PA-I, and TAP-1 proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the LSP-1, PA-I, and TAP-1 protein that maintain the ability to bind a LSP-1, PA-I, and TAP-1 ligand and/or modulate a LSP-1, PA-I, and TAP-1 activity. Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:2, 5, or 8, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein.

Non-functional allelic variants are naturally occurring amino acid sequence variants of the human LSP-1, PA-I, and TAP-1 protein that do not have the ability to either bind an ARP ligand and/or modulate a LSP-1, PA-I, and TAP-1 activity. Non-

functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion or premature truncation of the amino acid sequence of SEQ ID NO:2, 5, or 8 or a substitution, insertion or deletion in critical residues or critical regions.

The present invention further provides non-human orthologues of the LSP-1, PA-I, and TAP-1 protein. Orthologues of the LSP-1, PA-I, and TAP-1 protein are proteins that are isolated from non-human organisms and possess the same LSP-1, PA-I, and TAP-1 ligand binding and/or modulation of LSP-1, PA-I, and TAP-1 activities of the LSP-1, PA-I, and TAP-1 protein. Orthologues of the LSP-1, PA-I, and TAP-1 protein can readily be identified as including an amino acid sequence that is substantially homologous to SEQ ID NO:2, 5, or 8, as defined herein.

Moreover, nucleic acid molecules encoding LSP-1, PA-I, and TAP-1 proteins from other species, and which, thus, have a nucleotide sequence which differs from the sequence of SEQ ID NO:1, 3, 4, 6, 7, and 9 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the LSP-1, PA-I, and TAP-1 cDNAs of the invention can be isolated based on their homology to the LSP-1, PA-I, and TAP-1 nucleic acids disclosed herein using these cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9. In other embodiment, the nucleic acid is at least 30, 50, 100, 250 or 500 nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50°C, preferably at 55°C, more preferably at 60°C, and even more preferably at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a

"naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In addition to naturally-occurring allelic variants of the LSP-1, PA-I, and TAP-1 sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, thereby leading to changes in the amino acid sequence of the encoded LSP-1, PA-I, and TAP-1 protein, without altering the functional ability of the LSP-1, PA-I, and TAP-1 protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of LSP-1, PA-I, and TAP-1 (e.g., the sequence of SEQ ID NO:2, 5, or 8) without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues of PA-I that are conserved among the prolactin-growth hormone family members of this invention (as indicated by the alignment and comparison of the amino acid sequences of SEQ ID NOs:5 and 10 presented as Figure 5) are predicted to be essential in PA-I and, thus, are not likely to be amenable to alteration. For example, most proteins of the prolactin-growth hormone family, as well as the PA-I protein of the present invention, contain at least four cysteine residues among the cysteine rich domains (residues 101, 218, 235 and 244 of SEQ ID NO:5). Furthermore, amino acid residues that are conserved between LSP-1 protein and other proteins having Ig-like or Ig domains are not likely to be amenable to alteration. Moreover, amino acid residues of TAP-1 that are conserved among the family members of this invention (as indicated by an alignment and comparison of the amino acid sequences of SEQ ID NO:8 with sequences of TPO, e.g., human TPO shown in Figure 8) are predicted to be essential in TAP-1 and, thus, are not likely to be amenable to alteration. Identical or conserved amino acid sequences between TAP-1 and human TPO are indicated in the middle row between these sequences in Figure 8.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding LSP-1, PA-I, and TAP-1 proteins that contain changes in amino acid residues that are not essential for activity. Such LSP-1, PA-I, and TAP-1 proteins differ in amino acid sequence from SEQ ID NO:2, 5, or 8 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 60% homologous to the amino acid sequence of SEQ ID NO:2, 5, or 8. Preferably, the protein encoded by the nucleic acid molecule is at least about 70% homologous to SEQ ID NO:2, 5, or 8, more preferably at least about 80% homologous to SEQ ID NO:2, 5, or

8, even more preferably at least about 90% homologous to SEQ ID NO:2, 5, or 8, and most preferably at least about 95% homologous to SEQ ID NO:2, 5, or 8.

An isolated nucleic acid molecule encoding a LSP-1, PA-I, and TAP-1 protein homologous to the protein of SEQ ID NO:2, 5, or 8 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NO:1, 3, 4, 6, 7, or 9 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in LSP-1, PA-I, and TAP-1 is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a LSP-1, PA-I, and TAP-1 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for LSP-1, PA-I, and TAP-1 biological activity activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1, 3, 4, 6, 7, or 9, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant LSP-1 protein can be assayed for (1) the ability to modulate cellular signal transduction; (2) regulation of a cell involved in an inflammatory response; (3) homing of a cell having a LSP-1 protein on its cell surface from a first to a second anatomical location; and (4) the ability to modulate a cell involved in the immune response.

In another preferred embodiment, a mutant PA-I protein can be assayed for (1) the ability to form a complex with a cell-surface protein(s) or a ligand, e.g., a lipid or carbohydrate; (2) the ability to form a complex with a prolactin and/or growth hormone receptor. In yet another preferred embodiment, a mutant LSP-1, PA-I, and TAP-1 can be assayed for the ability to (1) regulate cellular growth; (2) regulate cellular

proliferation; (3) regulate angiogenesis; (4) regulate cellular differentiation; and (5) regulate cell survival.

In yet another preferred embodiment, a mutant TAP-1 protein can be assayed for (1) the ability to form protein:protein interactions with cell-surface proteins, e.g., a TAP-1 receptor, or biologically active portions thereof; (2) modulation of, e.g., activation or inhibition of, a cell-surface receptor; (3) modulation of cellular signal transduction.. In yet another preferred embodiment, a mutant TAP-1 can be assayed for the ability to (1) modulate cellular signal transduction; (2) regulate cellular proliferation; (3) regulate cellular differentiation; (4) regulate cell survival; (5) modulate a cell involved in the immune response; (6) regulation of maturation and/or differentiation of a hematopoietic stem cell; (7) modulation of megakaryocytopoiesis; (8) modulation of thrombopoiesis; (9) regulation of maturation and/or differentiation of a megakaryocyte into platelets; and (10) regulation of maturation and/or differentiation of erythroid progenitor cells into erythrocytes.

In addition to the nucleic acid molecules encoding LSP-1, PA-I, and TAP-1 proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire LSP-1, PA-I, and TAP-1 coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding LSP-1, PA-I, and TAP-1. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the coding region of LSP-1, PA-I, and TAP-1 corresponds to SEQ ID NO:3, 6, and 9, respectively). In another embodiment, the antisense nucleic acid molecule is antisense to a "non-coding region" of the coding strand of a nucleotide sequence encoding LSP-1, PA-I, and TAP-1. The term "non-coding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding LSP-1, PA-I, and TAP-1 disclosed herein (e.g., SEQ ID NO:3, 6, and 9, respectively), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of LSP-1, PA-I, and TAP-1 mRNA, but more preferably is an oligonucleotide which is

antisense to only a portion of the coding or noncoding region of LSP-1, PA-I, and TAP-1 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of LSP-1, PA-I, and TAP-1 mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5- oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a LSP-1, PA-I, and TAP-1 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the

invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave LSP-1, PA-I, and TAP-1 mRNA transcripts to thereby inhibit translation of LSP-1, PA-I, and TAP-1 mRNA. A ribozyme having specificity for a LSP-1, PA-I, and TAP-1-encoding nucleic acid can be designed based upon the nucleotide sequence of a LSP-1, PA-I, and TAP-1 cDNA disclosed herein (i.e., SEQ ID NO:1). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a LSP-1, PA-I, and TAP-1-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, LSP-1, PA-I, and TAP-1 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, LSP-1, PA-I, and TAP-1 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the LSP-1, PA-I, and TAP-1 (e.g., the LSP-1, PA-I, and TAP-1 promoter and/or enhancers) to form triple helical structures that prevent transcription of the LSP-1, PA-I, and TAP-1 gene in

target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. et al. (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

In preferred embodiments, the nucleic acids of LSP-1, PA-I, and TAP-1 can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *et al.* (1996) PNAS 93: 14670-675.

PNAs of LSP-1, PA-I, and TAP-1 can be used therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, e.g., inducing trascription or translation arrest or inhibiting replication. PNAs of LSP-1, PA-I, and TAP-1 can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as 'artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup B. (1996) *supra*); or as probes or primers for DNA sequence and hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

In another embodiment, PNAs of LSP-1, PA-I, and TAP-1 can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of LSP-1, PA-I, and TAP-1 can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNAse H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup B. (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup B. (1996) *supra* and Finn P.J. *et al.* (1996) *Nucleic Acids Research* 24 (17): 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, e.g., 5'-(4-methoxytrityl)amino-5'-deoxy-

thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acid Res.* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al., 1989, Proc. Natl. Acad. Sci. U.S.A. 86:6553-6556; Lemaitre et al., 1987, Proc. Natl. Acad. Sci. 84:648-652; PCT Publication No. W088/09810, published December 15, 1988) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134, published April 25, 1988). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, e.g., Krol et al., 1988, BioTechniques 6:958-976) or intercalating agents. (See, e.g., Zon, 1988, Pharm. Res. 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

II. Isolated LSP-1, PA-I, and TAP-1 Proteins and Anti-LSP-1, Anti-PA-I, and Anti-TAP-1 Antibodies

One aspect of the invention pertains to isolated LSP-1, PA-I, and TAP-1 proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-LSP-1, anti-PA-I, and anti-TAP-1 antibodies. In one embodiment, native LSP-1, PA-I, and TAP-1 proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, LSP-1, PA-I, and TAP-1 proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a LSP-1, PA-I, and TAP-1 protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the LSP-1, PA-I, and TAP-1 protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of LSP-1, PA-I, and TAP-1 protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of LSP-1, PA-I,

and TAP-1 protein having less than about 30% (by dry weight) of non-LSP-1, PA-I, and TAP-1 protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-LSP-1, PA-I, and TAP-1 protein, still more preferably less than about 10% of non-LSP-1, PA-I, and TAP-1 protein, and most preferably less than about 5% non-LSP-1, PA-I, and TAP-1 protein. When the LSP-1, PA-I, and TAP-1 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of LSP-1, PA-I, and TAP-1 protein in which the protein is separated from chemical precusors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of LSP-1, PA-I, and TAP-1 protein having less than about 30% (by dry weight) of chemical precursors or non-LSP-1, PA-I, and TAP-1 chemicals, more preferably less than about 20% chemical precursors or non-LSP-1, PA-I, and TAP-1 chemicals, still more preferably less than about 10% chemical precursors or non-LSP-1, PA-I, and TAP-1 chemicals, and most preferably less than about 5% chemical precursors or non-LSP-1, PA-I, and TAP-1 chemicals.

Biologically active portions of a LSP-1, PA-I, and TAP-1 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the LSP-1, PA-I, and TAP-1 protein, e.g., the amino acid sequence shown in SEQ ID NO:2, 5, or 8, which include less amino acids than the full length LSP-1, PA-I, and TAP-1 proteins, and exhibit at least one activity of a LSP-1, PA-I, and TAP-1 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the LSP-1, PA-I, and TAP-1 protein. A biologically active portion of a LSP-1, PA-I, and TAP-1 protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

In one embodiment, a biologically active portion of a LSP-1 protein comprises at least a transmembrane domain. In yet another embodiment, a biologically active portion of a LSP-1 protein comprises at least a signal sequence.

In an alternative embodiment, a biologically active portion of a LSP-1 protein comprises a LSP-1 amino acid sequence lacking a transmembrane domain. In yet another embodiment, a biologically active portion of a LSP-1 protein comprises a LSP-1 amino acid sequence lacking a transmembrane domain and a signal sequence. Such a preferred LSP-1 molecules are referred to as a "LSP-1 extracellular domains". For

example, preferred LSP-1 extracellular domains contain at least about amino acids 1-190 of SEQ ID NO:2.

In another embodiment, a biologically active portion of a PA-I protein comprises at least one cysteine rich domain, characteristic of the prolactin-growth hormone superfamily of proteins. In another embodiment, a biologically active portion of a PA-I, protein comprises at least a signal sequence.

In yet another embodiment, a biologically active portion of a TAP-1 protein comprises at least one serine-proline-threonine rich region of TAP-1.

In an alternative embodiment, a biologically active portion of a TAP-1 protein comprises at least a C-terminal unique domain of a TAP-1 protein.

It is to be understood that a preferred biologically active portion of a LSP-1, PA-I, and TAP-1 protein of the present invention may contain at least one of the above-identified structural domains. A more preferred biologically active portion of a LSP-1, PA-I, and TAP-1 protein may contain at least two of the above-identified structural domains. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native LSP-1, PA-I, and TAP-1 protein.

In a preferred embodiment, the LSP-1, PA-I, and TAP-1 protein has an amino acid sequence shown in SEQ ID NO:2, 5, or 8. In other embodiments, the LSP-1, PA-I, and TAP-1 protein is substantially homologous to SEQ ID NO:2, 5, or 8 and retains the functional activity of the protein of SEQ ID NO:2, 5, or 8 yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the LSP-1, PA-I, and TAP-1 protein is a protein which comprises an amino acid sequence at least about 60% homologous to the amino acid sequence of SEQ ID NO:2, 5, or 8 and retains the functional activity of the LSP-1, PA-I, and TAP-1 proteins of SEQ ID NO:2, 5, or 8. Preferably, the protein is at least about 70% homologous to SEQ ID NO:2, 5, or 8, more preferably at least about 80% homologous to SEQ ID NO:2, 5, or 8, even more preferably at least about 90% homologous to SEQ ID NO 2, and most preferably at least about 95% or more homologous to SEQ ID NO:2, 5, or 8. In a preferred embodiment the LSP-1 protein is at least 52%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more homologous to SEQ ID NO:2.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence

aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence (e.g., when aligning a second sequence to the LSP-1, PA-I, and TAP-1 amino acid sequence of SEQ ID NO:2, 5, or 8 having 177 amino acid residues, at least 80, preferably at least 100, more preferably at least 120, even more preferably at least 140, and even more preferably at least 150, 160 or 170 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to LSP-1, PA-I, and TAP-1 nucleic acid molecules of the invention.

BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to LSP-1, PA-I, and TAP-1 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

The invention also provides LSP-1, PA-I, and TAP-1 chimeric or fusion proteins. As used herein, a LSP-1, PA-I, and TAP-1 "chimeric protein" or "fusion protein" comprises a LSP-1, PA-I, and TAP-1 polypeptide operatively linked to a non-LSP-1, non-PA-I, and non-TAP-1 polypeptide. A "LSP-1, PA-I, and TAP-1 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to LSP-1, PA-I, and TAP-1, whereas a "non-LSP-1, non-PA-I, and non-TAP-1 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the LSP-1, PA-I, and TAP-1 protein, e.g., a protein which is different from the LSP-1, PA-I, and TAP-1 protein and which is derived from the same or a different organism. Within a LSP-1, PA-I, and TAP-1 fusion protein the LSP-1, PA-I, and TAP-1 polypeptide can correspond to all or a portion of a LSP-1, PA-I, and TAP-1 protein. In a preferred embodiment, a LSP-1, PA-I, and TAP-1 fusion protein comprises at least one biologically active portion of a LSP-1, PA-I, and TAP-1 protein. In another preferred embodiment, a LSP-1, PA-I, and TAP-1 fusion protein comprises at least two biologically active portions of a LSP-1, PA-I, and TAP-1 protein. In another preferred embodiment, a LSP-1, PA-I, and TAP-1 fusion protein comprises at least three biologically active portions of a LSP-1, PA-I, and TAP-1 protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the LSP-1, PA-I, and TAP-1 polypeptide and the non-LSP-1, non-PA-I, and non-TAP-1 polypeptide are fused in-frame to each other. The non-LSP-1, non-PA-I, and non-TAP-1 polypeptide can be fused to the N-terminus or C-terminus of the LSP-1, PA-I, and TAP-1 polypeptide. Such fusion proteins can be further utilized in screening assays for compounds which modulate LSP-1, PA-I, and TAP-1 activity (such assays are described in detail below).

In yet another embodiment, the fusion protein is a GST-LSP-1, GST-PA-I, and GST-TAP-1 fusion protein in which the LSP-1, PA-I, and TAP-1 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant LSP-1, PA-I, and TAP-1.

In another embodiment, the fusion protein is a LSP-1, PA-I, and TAP-1 protein containing a heterologous signal sequence at its N-terminus. For example, the native LSP-1, PA-I, and TAP-1 signal sequence (i.e, amino acids 1 to 20 of SEQ ID NO:2, or

amino acids 1 to 30 of SEQ ID NO:5) can be removed and replaced with a signal sequence from another protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of LSP-1, PA-I, and TAP-1 can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is a LSP-1, PA-I, and TAP-1-immunoglobulin fusion protein in which the LSP-1, PA-I, and TAP-1 sequences comprising primarily the cysteine rich domain are fused to sequences derived from a member of the immunoglobulin protein family. Methods for preparing such fusion proteins have been described in Capon, D.J. *et al.* (1989) *Nature* 337:525-531 and Capon U.S. Patents 5,116,964 and 5,428,130 [CD4-IgG1 constructs]; Linsley, P.S. *et al.* (1991) *J. Exp. Med.* 173:721-730 [a CD28-IgG1 construct and a B7-1-IgG1 construct]; and Linsley, P.S. *et al.* (1991) *J. Exp. Med.* 174:561-569 and U.S. Patent 5,434,131[a CTLA4-IgG1]). Such fusion proteins have proven useful for modulating receptor-ligand interactions.

The LSP-1, PA-I, and TAP-1-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a LSP-1, PA-I, and TAP-1 molecule and a protein which binds LSP-1, PA-I, and TAP-1 and exists on the surface of a cell (e.g., a LSP-1, PA-I, and TAP-1 receptor) to, thereby, suppress LSP-1, PA-I, and TAP-1-mediated signal transduction *in vivo*. The LSP-1, PA-I, and TAP-1-immunoglobulin fusion proteins can be used to affect the bioavailability of the LSP-1, PA-I, and TAP-1 molecule. Inhibition of the LSP-1, PA-I, and TAP-1 receptor/LSP-1, PA-I, and TAP-1 interaction may be useful therapeutically for both the treatment of LSP-1, PA-I, and TAP-1 associated disorders, e.g., proliferative and differentiative disorders, as well as modulating (e.g., promoting or inhibiting) angiogenesis. Moreover, the LSP-1, PA-I, and TAP-1-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-LSP-1, PA-I, and TAP-1 antibodies in a subject, to purify LSP-1, PA-I, and TAP-1 receptors and in screening assays to identify molecules which inhibit the interaction of LSP-1, PA-I, and TAP-1 with a LSP-1, PA-I, and TAP-1 receptor.

Preferably, a LSP-1, PA-I, and TAP-1 chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated

DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A LSP-1, PA-I, and TAP-1-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the LSP-1, PA-I, and TAP-1 protein.

The present invention also pertains to variants of the LSP-1, PA-I, and TAP-1 protein which function as either LSP-1, PA-I, and TAP-1 agonists (mimetics) or as LSP-1, PA-I, and TAP-1 antagonists. Variants of the LSP-1, PA-I, and TAP-1 protein can be generated by mutagenesis, e.g., discrete point mutation or truncation of the LSP-1, PA-I, and TAP-1 protein. An agonist of the LSP-1, PA-I, and TAP-1 protein can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the LSP-1, PA-I, and TAP-1 protein. An antagonist of the LSP-1, PA-I, and TAP-1 protein can inhibit one or more of the activities of the naturally occurring form of the LSP-1, PA-I, and TAP-1 protein by, for example, competitively binding to a LSP-1, PA-I, and TAP-1 receptor. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the LSP-1, PA-I, and TAP-1 proteins.

In one embodiment, variants of the LSP-1, PA-I, and TAP-1 protein which function as either LSP-1, PA-I, and TAP-1 agonists (mimetics) or as LSP-1, PA-I, and TAP-1 antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the LSP-1, PA-I, and TAP-1 protein for LSP-1, PA-I, and TAP-1 protein agonist or antagonist activity. In one embodiment, a variegated library of LSP-1, PA-I, and TAP-1 variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of LSP-1, PA-I, and TAP-1 variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential LSP-1, PA-I, and TAP-1 sequences is expressible as individual polypeptides or, alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of LSP-1, PA-I, and TAP-1 sequences therein. There are a variety of methods which can be used to produce libraries of potential LSP-1, PA-I, and TAP-1 variants from a degenerate oligonucleotide sequence. Chemical synthesis of a

degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential LSP-1, PA-I, and TAP-1 sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of the LSP-1, PA-I, and TAP-1 protein coding sequence can be used to generate a variegated population of LSP-1, PA-I, and TAP-1 fragments for screening and subsequent selection of variants of a LSP-1, PA-I, and TAP-1 protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a LSP-1, PA-I, and TAP-1 coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the LSP-1, PA-I, and TAP-1 protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of LSP-1, PA-I, and TAP-1 proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recrusive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify LSP-1, PA-I, and TAP-1 variants (Arkin and Yourvan (1992) *PNAS 89*:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated LSP-1 library. For example, a library of expression vectors can be transfected into a cell line which ordinarily responds to a particular ligand in a LSP-1-dependent manner. The transfected cells are then contacted with the ligand and the effect of expression of the

mutant on signaling by the ligand can be detected, e.g., by measuring any of a number of immune cell responses. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of ligand induction, and the individual clones further characterized.

In another embodiment, cell based assays can be exploited to analyze a variegated PA-I library. For example, a library of expression vectors can be transfected into a cell line which ordinarily responds to a PA-I molecule, e.g., a cell line derived from placental tissue. The transfected cells are then contacted with the PA-I mutant and the effect of expression of the mutant on cellular proliferation can be detected, e.g., by measuring a cellular proliferation-related parameter. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of cellular proliferation, and the individual clones further characterized.

In yet another embodiment, cell based assays can be exploited to analyze a variegated TAP-1 library. For example, a library of expression vectors containing the appropriate secretory signals can be transfected into a cell line and contacted with a HARP-responsive cell line. The effect of expression of the mutant on a TAP-1-responsive cell can be measured as, e.g., changes in signal transduction or by measuring cell proliferation, differentiation or survival. Plasmid DNA can then be recovered from the cells which score for modulation, e.g., inhibition, or potentiation, pf TAP-1 activity, and the individual clones further characterized.

An isolated LSP-1, PA-I, and TAP-1 protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind LSP-1, PA-I, and TAP-1 using standard techniques for polyclonal and monoclonal antibody preparation. The full-length LSP-1, PA-I, and TAP-1 protein can be used or, alternatively, the invention provides antigenic peptide fragments of LSP-1, PA-I, and TAP-1 for use as immunogens. The antigenic peptide of LSP-1, PA-I, and TAP-1 comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2, 5, or 8 and encompasses an epitope of LSP-1, PA-I, and TAP-1 such that an antibody raised against the peptide forms a specific immune complex with LSP-1, PA-I, and TAP-1. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of LSP-1, PA-I, and TAP-1 that are located on the surface of the protein, e.g., hydrophilic regions.

A LSP-1, PA-I, and TAP-1 immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example,

recombinantly expressed LSP-1, PA-I, and TAP-1 protein or a chemically synthesized LSP-1, PA-I, and TAP-1 polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic LSP-1, PA-I, and TAP-1 preparation induces a polyclonal anti-LSP-1, PA-I, and TAP-1 antibody response.

Accordingly, another aspect of the invention pertains to anti-LSP-1, PA-I, and TAP-1 antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as LSP-1, PA-I, and TAP-1. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind LSP-1, PA-I, and TAP-1. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of LSP-1, PA-I, and TAP-1. A monoclonal antibody composition thus typically displays a single binding affinity for a particular LSP-1, PA-I, and TAP-1 protein with which it immunoreacts.

Polyclonal anti-LSP-1, PA-I, and TAP-1 antibodies can be prepared as described above by immunizing a suitable subject with a LSP-1, PA-I, and TAP-1 immunogen. The anti-LSP-1, PA-I, and TAP-1 antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized LSP-1, PA-I, and TAP-1. If desired, the antibody molecules directed against LSP-1, PA-I, and TAP-1 can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the anti-LSP-1, PA-I, and TAP-1 antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown et al. (1981) *J. Immunol.* 127:539-46; Brown et al. (1980) *J. Biol. Chem* .255:4980-83; Yeh et al. (1976) *PNAS* 76:2927-31; and Yeh et al. (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor et al. (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological*

Analyses, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter et al. (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a LSP-1, PA-I, and TAP-1 immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds LSP-1, PA-I, and TAP-1.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-LSP-1, PA-I, and TAP-1 monoclonal antibody (see, e.g., G. Galfre et al. (1977) *Nature* 266:55052; Gefter et al. *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (e.g., a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, e.g., the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind LSP-1, PA-I, and TAP-1, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-LSP-1, PA-I, and TAP-1 antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with LSP-1, PA-I, and TAP-1 to thereby isolate immunoglobulin library members that bind LSP-1, PA-I, and TAP-1. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents

particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner et al. U.S. Patent No. 5,223,409; Kang et al. PCT International Publication No. WO 92/18619; Dower et al. PCT International Publication No. WO 91/17271; Winter et al. PCT International Publication WO 92/20791; Markland et al. PCT International Publication No. WO 92/15679; Breitling et al. PCT International Publication WO 93/01288; McCafferty et al. PCT International Publication No. WO 92/01047; Garrard et al. PCT International Publication No. WO 92/09690; Ladner et al. PCT International Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J* 12:725-734; Hawkins et al. (1992) *J. Mol. Biol.* 226:889-896; Clarkson et al. (1991) *Nature* 352:624-628; Gram et al. (1992) *PNAS* 89:3576-3580; Garrad et al. (1991) *Bio/Technology* 9:1373-1377; Hoogenboom et al. (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas et al. (1991) *PNAS* 88:7978-7982; and McCafferty et al. *Nature* (1990) 348:552-554.

Additionally, recombinant anti-LSP-1, PA-I, and TAP-1 antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson et al. International Application No. PCT/US86/02269; Akira, et al. European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al. European Patent Application 173,494; Neuberger et al. PCT International Publication No. WO 86/01533; Cabilly et al. U.S. Patent No. 4,816,567; Cabilly et al. European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *PNAS* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *PNAS* 84:214-218; Nishimura et al. (1987) *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al. (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison, S. L. (1985) *Science* 229:1202-1207; Oi et al. (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; and Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

An anti-LSP-1, PA-I, and TAP-1 antibody (e.g., monoclonal antibody) can be used to isolate LSP-1, PA-I, and TAP-1 by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-LSP-1, PA-I, and TAP-1 antibody can facilitate the purification of natural LSP-1, PA-I, and TAP-1 from cells and of recombinantly produced LSP-1, PA-I, and TAP-1 expressed in host cells. Moreover, an anti-LSP-1, PA-I, and TAP-1 antibody can be used to detect LSP-1, PA-I, and TAP-1

protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the LSP-1, PA-I, and TAP-1 protein. Anti-LSP-1, PA-I, and TAP-1 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding LSP-1, PA-I, and TAP-1 (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective

retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., LSP-1, PA-I, and TAP-1 proteins, mutant forms of LSP-1, PA-I, and TAP-1, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of LSP-1, PA-I, and TAP-1 in prokaryotic or eukaryotic cells. For example, LSP-1, PA-I, and TAP-1 can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promotors directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein;

2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

In a preferred embodiment, the coding sequence of human LSP-1, PA-I, and TAP-1 (i.e., encompassing amino acids 1 to 253) is cloned into a pCD5 expression vector to create a vector encoding a LSP-1, PA-I, and TAP-1-Ig fusion protein. In an alternative preferred embodiment, the coding sequence of a form of human LSP-1, PA-I, and TAP-1 lacking the signal sequence is cloned into a pPicZ expression vector (InVitrogen) downstream and in frame with a yeast-derived signal sequence. In yet another preferred embodiment, the coding sequence of human LSP-1, PA-I, and TAP-1 (e.g., the sequence shown in SEQ ID NO:3, 6, or 9) is cloned into a retroviral expression vector, pWZLBlastEC. The fusion proteins can be purified utilizing methods well known in the art of protein purification. Purified fusion proteins can be utilized in LSP-1, PA-I, and TAP-1 activity assays, in LSP-1, PA-I, and TAP-1 receptor binding (e.g. direct assays or competitive assays described in detail below), to generate antibodies specific for LSP-1, PA-I, and TAP-1 proteins, as examples. In a preferred embodiment, a LSP-1, PA-I, and TAP-1 fusion expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g six (6) weeks).

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al., (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the LSP-1, PA-I, and TAP-1 expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari, et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (InVitrogen Corp, San Diego, CA).

Alternatively, LSP-1, PA-I, and TAP-1 can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and

Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to LSP-1, PA-I, and TAP-1 mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, LSP-1, PA-I, and TAP-1 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells, such as Chinese hamster ovary cells (CHO) or COS cells. Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding LSP-1, PA-I, and TAP-1 or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) LSP-1, PA-I, and TAP-1 protein. Accordingly, the invention further provides methods for producing LSP-1, PA-I, and TAP-1 protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding LSP-1, PA-I, and TAP-1 has been introduced) in a suitable medium such that LSP-1, PA-I, and TAP-1 protein is produced. In another embodiment, the method further comprises isolating LSP-1, PA-I, and TAP-1 from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which LSP-1, PA-I, and TAP-1-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous LSP-1, PA-I, and TAP-1 sequences have been introduced into their genome or homologous recombinant animals in which endogenous LSP-1, PA-I, and TAP-1 sequences have been altered. Such animals are useful for studying the function and/or activity of LSP-1, PA-I, and TAP-1 and for identifying and/or evaluating

modulators of LSP-1, PA-I, and TAP-1 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous LSP-1, PA-I, and TAP-1 gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing LSP-1, PA-I, and TAP-1-encoding nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The murine LSP-1, PA-I, and TAP-1 cDNA sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9 can be introduced as a transgene into the genome of a non-murine animal. Alternatively, a nonmurine homologue of the murine LSP-1, PA-I, and TAP-1 gene, such as a human LSP-1, PA-I, and TAP-1 gene, can be isolated based on hybridization to the murine LSP-1, PA-I, and TAP-1 cDNA (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the LSP-1, PA-I, and TAP-1 transgene to direct expression of LSP-1, PA-I, and TAP-1 protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Patent No. 4,873,191 by Wagner et al. and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the LSP-1, PA-I, and TAP-1 transgene in its genome and/or expression of LSP-1, PA-I, and TAP-1 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding LSP-1, PA-I, and TAP-1 can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a LSP-1, PA-I, and TAP-1 gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the LSP-1, PA-I, and TAP-1 gene. The LSP-1, PA-I, and TAP-1 gene can be a human gene, but more preferably, is a non-human homologue of a human LSP-1, PA-I, and TAP-1 gene. For example, a mouse LSP-1, PA-I, and TAP-1 gene of SEQ ID NO:1 can be used to construct a homologous recombination vector suitable for altering an endogenous LSP-1, PA-I, and TAP-1 gene in the mouse genome. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous LSP-1, PA-I, and TAP-1 gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous LSP-1, PA-I, and TAP-1 gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous LSP-1, PA-I, and TAP-1 protein). In the homologous recombination vector, the altered portion of the LSP-1, PA-I, and TAP-1 gene is flanked at its 5' and 3' ends by additional nucleic acid of the LSP-1, PA-I, and TAP-1 gene to allow for homologous recombination to occur between the exogenous LSP-1, PA-I, and TAP-1 gene carried by the vector and an endogenous LSP-1, PA-I, and TAP-1 gene in an embryonic stem cell. The additional flanking LSP-1, PA-I, and TAP-1 nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K.R. and Capecchi, M. R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced LSP-1, PA-I, and TAP-1 gene has homologously recombined with the endogenous LSP-1, PA-I, and TAP-1 gene are selected (see e.g., Li, E. et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec et

al.; WO 91/01140 by Smithies et al.; WO 92/0968 by Zijlstra et al.; and WO 93/04169 by Berns et al.

In another embodiment, transgenic non-humans animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso et al. (1992) *PNAS* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. et al. (1997) *Nature* 385:810-813. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The recontructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

IV. Pharmaceutical Compositions

The LSP-1, PA-I, and TAP-1 nucleic acid molecules, LSP-1, PA-I, and TAP-1 proteins, and anti-LSP-1, anti-PA-I, and anti-TAP-1 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a LSP-1, PA-I, and TAP-1 protein or anti-LSP-1, PA-I, and TAP-1 antibody) in the required amount in an appropriate solvent with one or a combination of

ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled

release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine

useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see e.g., Chen et al. (1994) *PNAS* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: 1) screening assays; 2) predictive assays (e.g., diagnostic assays and pharmacogenetics); 3) prognostic assays; 3) monitoring clinical trials; and 4) methods of treatment (e.g., therapeutic and prophylactic).

As described herein, a LSP-1 protein of the invention has one or more of the following activities: (i) interaction of a LSP-1 protein on the cell surface with a second non-LSP-1 protein molecule on the surface of the same cell; (ii) interaction of a LSP-1 protein on the cell surface with a second non-LSP-1 protein molecule on the surface of a different cell; (iii) complex formation between a soluble LSP-1 protein and a cognate ligand; (iv) complex formation between a membrane-bound LSP-1 protein and a cytokine; (v) interaction of a LSP-1 protein with an intracellular protein via a second protein on the cell surface and can thus be used for (i) modulation of cellular signal transduction; (ii) regulation of a cell involved in an inflammatory response; (iii) homing of a cell having a LSP-1 protein on its cell surface from a first to a second anatomical location; and (iv) modulation of a cell involved in the immune response, either *in vitro* or *in vivo*. The isolated nucleic acid molecules of the invention can be used, for example, to express LSP-1 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect LSP-1 mRNA (e.g., in a biological sample) or a genetic lesion in a LSP-1 gene, and to modulate LSP-1 activity, as described further below. In addition, the LSP-1 proteins can be used to screen drugs or compounds which

modulate the LSP-1 activity as well as to treat disorders characterized by insufficient or excessive production of LSP-1 protein or production of LSP-1 protein forms which have decreased or aberrant activity compared to LSP-1 wild type protein (e.g., inflammatory diseases such as arthritis and immune response disorders). Moreover, soluble forms of the LSP-1 protein can be used to bind ligands of membrane-bound LSP-1 and influence bioavailability. In addition, the anti-LSP-1 antibodies of the invention can be used to detect and isolate LSP-1 proteins and modulate LSP-1 activity.

As further described herein, a PA-I protein of the invention has one or more of the following activities: (i) formation of a complex with a cell-surface protein(s) or a ligand, e.g., a lipid or carbohydrate; (ii) formation of a complex with a prolactin and/or growth hormone receptor; (iii) regulation of cellular growth; (iv) regulation of cellular proliferation; (v) regulation of angiogenesis; (vi) regulation of cellular differentiation; and (vii) regulation of cell survival, and can thus be used to (i) modulate complex formation with a cell-surface protein(s) or a ligand, e.g., a lipid or carbohydrate; (ii) modulate complex formation with a prolactin and/or growth hormone receptor; (iii) regulate cellular growth; (iv) regulate cellular proliferation; (v) regulate angiogenesis; (vi) regulate cellular differentiation; and (vii) regulate cell survival, either *in vitro* or *in vivo*. The isolated nucleic acid molecules of the invention can be used, for example, to express PA-I protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect PA-I mRNA (e.g., in a biological sample) or a genetic lesion in a PA-I gene, and to modulate PA-I activity, as described further below. In addition, the PA-I proteins can be used to screen drugs or compounds which modulate the PA-I activity as well as to treat disorders characterized by insufficient or excessive production of PA-I protein or production of PA-I protein forms which have decreased or aberrant activity compared to PA-I wild type protein (e.g., proliferative disorders such as cancer or angiogenesis related disorders). Moreover, soluble forms of PA-I antagonists can be used to bind membrane-bound PA-I receptors and influence bioavailability. In addition, the anti-PA-I antibodies of the invention can be used to detect and isolate PA-I proteins and modulate PA-I activity.

Furthermore, as described herein, a TAP-1 protein of the invention has the following activities: (i) interaction, e.g., binding to, a cell-surface receptor, e.g., a hematopoietic-cell surface receptor; (ii) modulation of, e.g., activation or inhibition of, a cell-surface receptor (iii) modulation of cellular signal transduction and can thus be used to (i) regulate cellular proliferation; (ii) regulate cellular differentiation; (iii) regulate cell survival; (iv) modulate a cell involved in the immune response (v) regulate maturation and/or differentiation of a hematopoietic stem cell; (vi) modulate megakaryocytopoiesis; (vii) modulate thrombopoiesis; (viii) regulate maturation and/or differentiation of a

megakaryocyte into platelets; and (ix) regulate maturation and/or differentiation of erythroid progenitor cells into erythrocytes, either *in vitro* or *in vivo*. The isolated nucleic acid molecules of the invention can be used to express TAP-1 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect TAP-1 mRNA (e.g., in a biological sample) or a genetic lesion in a TAP-1 gene, and to modulate TAP-1 activity, as described further below. In addition, the TAP-1 proteins can be used to screen drugs or compounds which modulate the TAP-1 activity as well as to treat disorders characterized by insufficient or excessive production of TAP-1 protein or production of TAP-1 protein forms which have decreased or abherrent activity compared to TAP-1 wild type protein (e.g. hematopoietic disorders disorders such as thrombocytopenia or anemia). Moreover, soluble forms of the TAP-1 protein can be used to bind ligands of membrane-bound TAP-1 and influence bioavailability. In addition, the anti-TAP-1 antibodies of the invention can be used to detect and isolate TAP-1 proteins and modulate TAP-1 activity.

A. Screening Assays:

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which have a LSP-1, PA-I, and TAP-1 agonist or antagonist activity or have a stimulatory or inhibitory effect on, for example, LSP-1, PA-I, and TAP-1 expression or LSP-1, PA-I, and TAP-1 activity.

In one embodiment, the invention provides assays for screening candidate or test compounds which have a LSP-1, PA-I, and TAP-1 agonist or antagonist activity. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al. (1994). *J. Med. Chem.* 37:2678; Cho et al. (1993) *Science* 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop et al. (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull et al. (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990)*Science* 249:404-406); (Cwirla et al. (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses or responds to a LSP-1, PA-I, and TAP-1 protein, or a biologically active portion thereof, is contacted with a test compound and the ability of the test compound to inhibit or stimulate the biological activity of a LSP-1, PA-I, and TAP-1 protein is determined. The cell, for example, can be of a mammalian origin or can be a yeast cell. Determining the ability of the test compound to inhibit or stimulate the biological activity of a LSP-1, PA-I, and TAP-1 protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to a LSP-1, PA-I, and TAP-1 receptor can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

In another embodiment, an assay is a cell-based assay comprising contacting a cell which expresses or responds to a LSP-1, PA-I, and TAP-1 protein or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (e.g. stimulate or inhibit) the activity of the LSP-1, PA-I, and TAP-1 protein or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of LSP-1, PA-I, and TAP-1 or a biologically active portion thereof can be accomplished, for example, by determining the ability of the LSP-1, PA-I, and TAP-1 protein to bind to or interact with a LSP-1, PA-I, and TAP-1 target molecule. As used herein, a "target molecule" is a molecule with which a LSP-1, PA-I, and TAP-1 protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a LSP-1, PA-I, and TAP-1 protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A LSP-1, PA-I, and TAP-1 target molecule can be a non-LSP-1, PA-I, and TAP-1 molecule. In one embodiment, a LSP-1, PA-I, and TAP-1 target molecule is a component of a signal

transduction pathway which facilitates transduction of an extracellular signal (e.g. a signal generated by binding of a LSP-1, PA-I, and TAP-1 molecule to a membrane-bound receptor) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein which has catalytic activity or a substrate for a catalytic activity of the LSP-1, PA-I, and TAP-1 receptor.

Determining the ability of the LSP-1, PA-I, and TAP-1 protein to bind to or interact with a LSP-1, PA-I, and TAP-1 target molecule can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the LSP-1, PA-I, and TAP-1 protein to bind to or interact with a LSP-1, PA-I, and TAP-1 target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (i.e. intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a LSP-1, PA-I, and TAP-1-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g. luciferase), or detecting a cellular response, for example, cellular differentiation, or cell proliferation.

In another embodiment, the assay is a cell-free assay in which a LSP-1, PA-I, and TAP-1 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the LSP-1, PA-I, and TAP-1 protein or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of LSP-1, PA-I, and TAP-1 can be accomplished, for example, by determining the ability of the LSP-1, PA-I, and TAP-1 protein to bind to a LSP-1, PA-I, and TAP-1 target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of LSP-1, PA-I, and TAP-1 can be accomplished by determining the ability of the LSP-1, PA-I, and TAP-1 protein to further modulate a LSP-1, PA-I, and TAP-1 target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In any of the above disclosed assay methods of the present invention, it may be desirable to immobilize either LSP-1, PA-I, and TAP-1 or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to LSP-1, PA-I, and TAP-1, or interaction of LSP-1, PA-I, and TAP-1 with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such

vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ LSP-1, PA-I, and TAP-1 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or LSP-1, PA-I, and TAP-1 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of LSP-1, PA-I, and TAP-1 binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either LSP-1, PA-I, and TAP-1 or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated LSP-1, PA-I, and TAP-1 or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with LSP-1, PA-I, and TAP-1 or target molecules but which do not interfere with binding of the LSP-1, PA-I, and TAP-1 protein to its target molecule can be derivatized to the wells of the plate, and unbound target or LSP-1, PA-I, and TAP-1 trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the LSP-1, PA-I, and TAP-1 or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the LSP-1, PA-I, and TAP-1 or target molecule.

In another embodiment, modulators of LSP-1, PA-I, and TAP-1 expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of LSP-1, PA-I, and TAP-1 mRNA or protein in the cell is determined. The level of expression of LSP-1, PA-I, and TAP-1 mRNA or protein in the presence of the candidate compound is compared to the level of expression of LSP-1, PA-I, and TAP-1 mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of LSP-1, PA-I, and TAP-1 expression based on this comparison. For example, when expression of LSP-1, PA-I, and TAP-1 mRNA or

protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of LSP-1, PA-I, and TAP-1 mRNA or protein expression. Alternatively, when expression of LSP-1, PA-I, and TAP-1 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of LSP-1, PA-I, and TAP-1 mRNA or protein expression. The level of LSP-1, PA-I, and TAP-1 mRNA or protein expression in the cells can be determined by methods described herein for detecting LSP-1, PA-I, and TAP-1 mRNA or protein.

In yet another aspect of the invention, the LSP-1, PA-I, and TAP-1 proteins can be used as "bait proteins" in a two-hybrid assay or in a three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with LSP-1, PA-I, and TAP-1 ("LSP-1-binding proteins", "PA-I-binding proteins", and "TAP-1-binding proteins" or "LSP-1-bp", "PA-I-bp", and "TAP-1-bp") and modulate LSP-1, PA-I, and TAP-1 activity. Such LSP-1-, PA-I-, and TAP-1-binding proteins are also likely to be involved in the propagation of signals by the LSP-1, PA-I, and TAP-1 proteins as, for example, upstream or downstream elements of the LSP-1, PA-I, and TAP-1 pathway, e.g., a LSP-1, PA-I, and TAP-1 receptor.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for LSP-1, PA-I, and TAP-1 is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a LSP-1, PA-I, and TAP-1-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with LSP-1, PA-I, and TAP-1.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

B. **Predictive Medicine:**

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenetics, and monitoring clinical trails are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining LSP-1, PA-I, and TAP-1 protein and/or nucleic acid expression as well as LSP-1, PA-I, and TAP-1 activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant LSP-1, PA-I, and TAP-1 expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with LSP-1, PA-I, and TAP-1 protein, nucleic acid expression or activity. For example, mutations in a LSP-1, PA-I, and TAP-1 gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby phophylactically treat an individual prior to the onset of a disorder characterized by or associated with LSP-1, PA-I, and TAP-1 protein, nucleic acid expression or activity.

Another aspect of the invention provides methods for determining LSP-1, PA-I, and TAP-1 protein, nucleic acid expression or LSP-1, PA-I, and TAP-1 activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenetics). Pharmacogenetics allows for the selection of agents (e.g., drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (e.g., the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (e.g. drugs, compounds) on the expression or activity of LSP-1, PA-I, and TAP-1 in clinical trials. These and other agents are described in further detail in the following sections.

1. **Diagnostic Assays**

An exemplary method for detecting the presence or absence of LSP-1, PA-I, and TAP-1 in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting LSP-1, PA-I, and TAP-1 protein or nucleic acid, (e.g., mRNA, genomic DNA) that encodes LSP-1, PA-I, and TAP-1 protein such that the presence of LSP-1, PA-I, and TAP-1 is detected in the biological sample.

A preferred agent for detecting LSP-1, PA-I, and TAP-1 mRNA is a labeled nucleic acid probe capable of hybridizing to LSP-1, PA-I, and TAP-1 mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length LSP-1, PA-I, and TAP-1 nucleic acid, such as the nucleic acid of SEQ ID NO:1, 3, 4, 6, 7, or 9, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to LSP-1, PA-I, and TAP-1 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting LSP-1, PA-I, and TAP-1 protein is an antibody capable of binding to LSP-1, PA-I, and TAP-1 protein (preferably an antibody with a detectable label). Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect LSP-1, PA-I, and TAP-1 mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of LSP-1, PA-I, and TAP-1 mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of LSP-1, PA-I, and TAP-1 protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of LSP-1, PA-I, and TAP-1 genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of LSP-1, PA-I, and TAP-1 protein include introducing into a subject a labeled anti-LSP-1, PA-I, and TAP-1 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatiavely, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood lymphocyte sample isolated by conventional means from a subject.

In another emodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting LSP-1, PA-I, and TAP-1 protein, mRNA, or genomic DNA, such that the presence of LSP-1, PA-I, and TAP-1 protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of LSP-1, PA-I, and TAP-1 protein, mRNA or genomic DNA in the control sample with the presence of LSP-1, PA-I, and TAP-1 protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of LSP-1, PA-I, and TAP-1 in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting LSP-1, PA-I, and TAP-1 protein or mRNA in a biological sample; means for determining the amount of LSP-1, PA-I, and TAP-1 in the sample; and means for comparing the amount of LSP-1, PA-I, and TAP-1 in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect LSP-1, PA-I, and TAP-1 protein or nucleic acid.

2. Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant LSP-1, PA-I, and TAP-1 expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with LSP-1, PA-I, and TAP-1 protein, nucleic acid expression or activity such as an inflammatory or immune disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing an immune disease or disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant LSP-1, PA-I, and TAP-1 expression or activity in which a test sample is obtained from a subject and LSP-1, PA-I, and TAP-1 protein or nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of LSP-1, PA-I, and TAP-1 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant LSP-1, PA-I, and TAP-1 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate)

to treat a disease or disorder associated with aberrant LSP-1, PA-I, and TAP-1 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder, such as inflammatory disease or an immune system disease. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant LSP-1, PA-I, and TAP-1 expression or activity in which a test sample is obtained and LSP-1, PA-I, and TAP-1 protein or nucleic acid is detected (e.g. wherein the presence of LSP-1, PA-I, and TAP-1 protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant LSP-1, PA-I, and TAP-1 expression or activity.)

The methods of the invention can also be used to detect genetic lesions in a LSP-1, PA-I, and TAP-1 gene, thereby determining if a subject with the gene, which is lesioned, is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a LSP-1, PA-I, and TAP-1-protein, or the mis-expression of the LSP-1, PA-I, and TAP-1 gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of (1) a deletion of one or more nucleotides from a LSP-1, PA-I, and TAP-1 gene; (2) an addition of one or more nucleotides to a LSP-1, PA-I, and TAP-1 gene; (3) a substitution of one or more nucleotides of a LSP-1, PA-I, and TAP-1 gene, (4) a chromosomal rearrangement of a LSP-1, PA-I, and TAP-1 gene; (5) an alteration in the level of a messenger RNA transcript of a LSP-1, PA-I, and TAP-1 gene, (6) aberrant modification of a LSP-1, PA-I, and TAP-1 gene, such as of the methylation pattern of the genomic DNA, (7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a LSP-1, PA-I, and TAP-1 gene, (8) a non-wild type level of a LSP-1, PA-I, and TAP-1-protein, (9) allelic loss of a LSP-1, PA-I, and TAP-1 gene, and (10) inappropriate post-translational modification of a LSP-1, PA-I, and TAP-1-protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a LSP-1, PA-I, and TAP-1 gene.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1994) *PNAS* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the LSP-1, PA-I, and TAP-1-gene (see Abravaya et al. (1995) *Nucleic Acids Res* .23:675-682). This method can include the steps of collecting

a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a LSP-1, PA-I, and TAP-1 gene under conditions such that hybridization and amplification of the LSP-1, PA-I, and TAP-1-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.*, 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.*, 1989, Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. et all, 1988, Bio/Technology 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a LSP-1, PA-I, and TAP-1 gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in LSP-1, PA-I, and TAP-1 can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in LSP-1, PA-I, and TAP-1 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential ovelapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using

smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the LSP-1, PA-I, and TAP-1 gene and detect mutations by comparing the sequence of the sample LSP-1, PA-I, and TAP-1 with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) *PNAS* 74:560) or Sanger ((1977) *PNAS* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen et al. (1996) *Adv. Chromatogr.* 36:127-162; and Griffin et al. (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the LSP-1, PA-I, and TAP-1 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes or heteroduplexes (Myers et al. (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type LSP-1, PA-I, and TAP-1 sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton et al (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba et al (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in LSP-1, PA-I, and TAP-1 cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994)

Carcinogenesis 15:1657-1662). According to an exemplary embodiment, a probe based on a LSP-1, PA-I, and TAP-1 sequence, e.g., a wild-type LSP-1, PA-I, and TAP-1 sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in LSP-1, PA-I, and TAP-1 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; and Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control LSP-1, PA-I, and TAP-1 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) *Trends Genet* 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) *Nature* 324:163); Saiki et al. (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different

mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al. (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a LSP-1, PA-I, and TAP-1 gene.

Furthermore, any cell type or tissue, preferably peripheral blood lymphocytes, in which LSP-1, PA-I, and TAP-1 is expressed may be ulilized in the prognostic assays described herein.

3. Pharmacogenetics

Agents, or modulators which have a stimulatory or inhibitory effect on LSP-1, PA-I, and TAP-1 activity (e.g., LSP-1, PA-I, and TAP-1 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders associated with aberrant LSP-1, PA-I, and TAP-1 activity, e.g., proliferative disorders. In conjunction with such treatment, the pharmacogenetics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound, e.g., a drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenetics of the individual permit the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based

on a consideration of the individual's genotype. Such pharmacogenetics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of LSP-1, PA-I, and TAP-1 protein, expression of LSP-1, PA-I, and TAP-1 nucleic acid, or mutation content of LSP-1, PA-I, and TAP-1 genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenetics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See e.g., Eichelbaum, M., *Clin Exp Pharmacol Physiol*, 1996, 23(10-11) :983-985 and Linder, M.W., *Clin Chem*,1997, 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of LSP-1, PA-I, and TAP-1 protein, expression of LSP-1, PA-I, and TAP-1 nucleic acid, or mutation content of LSP-1, PA-I, and TAP-1 genes in an

individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a LSP-1, PA-I, and TAP-1 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

4. **Monitoring of Effects During Clinical Trials**

Monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of LSP-1, PA-I, and TAP-1 (e.g., the ability to modulate inflammation, immune responsiveness, or cellular homing) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay to increase LSP-1, PA-I, and TAP-1 gene expression, protein levels, or upregulate LSP-1, PA-I, and TAP-1 activity, can be monitored in clinical trails of subjectes exhibiting decreased LSP-1, PA-I, and TAP-1 gene expression, protein levels, or downregulated LSP-1, PA-I, and TAP-1 activity. Alternatively, the effectiveness of a compounds determined by a screening assay to decrease LSP-1, PA-I, and TAP-1 gene expression, protein levels, or downregulate LSP-1, PA-I, and TAP-1 activity, can be monitored in clinical trails of subjectes exhibiting increased LSP-1, PA-I, and TAP-1 gene expression, protein levels, or upregulated LSP-1, PA-I, and TAP-1 activity. In such clinical trials, the expression of LSP-1, PA-I, and TAP-1 and, preferably, other genes that have been implicated in a metabolic disorder as described herein, can be used as a "read out" or as a marker of the metabolic state of a particular cell.

For example, and not by way of limitation, genes, including LSP-1, PA-I, and TAP-1, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) which modulates angiogenesis, cellular proliferation, or immune responses (e.g., identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on metabolic disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of LSP-1, PA-I, and TAP-1 and other genes implicated in the metabolic disorder. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of LSP-1, PA-I, and TAP-1 or other gene. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the

cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g. an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a LSP-1, PA-I, and TAP-1 protein, mRNA, or genomic DNA in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the LSP-1, PA-I, and TAP-1 protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the LSP-1, PA-I, and TAP-1 protein, mRNA, or genomic DNA in the pre-administration sample with the LSP-1, PA-I, and TAP-1 protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of LSP-1, PA-I, and TAP-1 to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of LSP-1, PA-I, and TAP-1 to lower levels than detected, i.e., to decrease the effectiveness of the agent.

C. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk for (or succeptible to) a disorder or having a disorder associated with aberrant LSP-1, PA-I, and TAP-1 expression or activity.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant LSP-1, PA-I, and TAP-1 expression or activity, by administering to the subject an agent which modulates LSP-1, PA-I, and TAP-1 expression or at least one LSP-1, PA-I, and TAP-1 activity. Subjects at risk for a disease which is caused or contributed to by aberrant LSP-1, PA-I, and TAP-1 expression or activity can be identified by, for example, any or a combination of the diagnostic or prognostic assays as described herein. Administeration of a phophylactic agent can occur prior to the manifestation of symptoms characteristic of the LSP-1, PA-I, and TAP-1 aberrancy such that a disease or disorder is prevented or, alternatively,

delayed in its progression. Depending on the type of LSP-1, PA-I, and TAP-1 aberrancy, for example, a LSP-1, PA-I, and TAP-1 agonist or LSP-1, PA-I, and TAP-1 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the present invention are further described in the following subsection.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating LSP-1, PA-I, and TAP-1 expression or activity for therapeutic purposes. Another aspect of the invention pertains to methods of modulating LSP-1, PA-I, and TAP-1 expression or activity associated with a cell for therapeutic purposes. LSP-1, PA-I, and TAP-1 activity "associated with a cell" is intended to include one or more of the activites of LSP-1, PA-I, and TAP-1 protein within the cell, secreted by the cell and in the extracellular milieu surrounding the cell. The modulatory method of the invention involves contacting the cell with an agent that modulates one or more of the activites of LSP-1, PA-I, and TAP-1 protein activity associated with the cell. An agent that modulates LSP-1, PA-I, and TAP-1 protein activity can be an agent as described herein such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a LSP-1, PA-I, and TAP-1 protein, a peptide, a LSP-1, PA-I, and TAP-1 peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more LSP-1, PA-I, and TAP-1 protein activity. Examples of such stimulatory agents include active LSP-1, PA-I, and TAP-1 protein and a nucleic acid molecule encoding LSP-1, PA-I, and TAP-1 that has been introduced into the cell. In another embodiment, the agent inhibits one or more LSP-1, PA-I, and TAP-1 protein activity. Examples of such inhibitory agents include antisense LSP-1, PA-I, and TAP-1 nucleic acid molecules and anti-LSP-1, PA-I, and TAP-1 antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a LSP-1, PA-I, and TAP-1 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents, that modulates (e.g., upregulates or downregulates) LSP-1, PA-I, and TAP-1 expression or activity. In another embodiment, the method involves administering a LSP-1, PA-I, and TAP-1 protein or nucleic acid molecule as therapy to compensate for reduced or aberrant LSP-1, PA-I, and TAP-1 expression or activity.

Stimulation of LSP-1, PA-I, and TAP-1 activity is desirable in situations in which LSP-1, PA-I, and TAP-1 is abnormally downregulated and/or in which increased LSP-1, PA-I, and TAP-1 activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation (e.g., cancer). Another example of such a situation is where the subject has an angiogenesis disorder. One example of such a situation is where a subject has a disorder characterized by aberrant immune responsiveness. Another example of such a situation is where the subject has a inflammatory disease (e.g., arthritis).

In one embodiment, a method for modulating, e.g., stimulating or inhibiting, the proliferation, maturation, differentiation or survival of a stem cell, e.g., a hematopoietic stem cell, is provided. The term "stem cell" refers to an undifferentiated cell which is capable of self-renewal, i.e., proliferation to give rise to more stem cells, and may give rise to lineage committed progenitors which are capable of differentiation and expansion into a specific lineage. In a preferred embodiment, the term "stem cell" refers to a generalized mother cell whose descendants (progeny) specialize, often in different directions, by differentiation, e.g., by acquiring completely individual characters, as occurs in progressive diversification of embryonic cells and tissues. As used herein, the term "stem cells" refers generally to both embryonic and hematopoietic stem cells from mammalian origin, e.g., human.

For example, the present invention provides a method of differentiating hematopoietic cells can be differentiated into mature cell of erythroid, lymphoid or myeloid lineages, e.g., by potentiating or disrupting the biological activities of a TAP-1 protein. As used herein, the term "hematopoietic stem cell" (HSC) means a population of cells capable of both self-renewal and differentiation into all defined hematopoietic lineages, i.e., myeloid, lymphoid or erythroid lineages. HSCs can ultimately differentiate into hematopoietic cells, including without limitation, common lymphoid progenitor cells, T cells (e.g., helper, cytotoxic, and suppressor cells), B cells, plasma cells, natural killer cells, common myeloid progenitor cells, monocytes, macrophages, mast cells, leukocytes, basophils, neutrophils, eosinophils, megakaryocytes, platelets, and erythroids. Preferably, the hematopoietic cells are selected from megakaryocytes, platelets, and erythroids.

The the present invention further provides methods for treating hematopoietic diseases. Examples of hematopoietic diseases include thrombocytopenia associated with bone marrow hypoplasia (e.g., aplastic anemia following radio- or chemotherapy or bone marrow transplant), immune thrombocytopenia (HIV and non-HIV induced thrombocytopenia), disorders such as intravascular coagulation, myeloproliferative

thrombocytotic disorders, inflammatory thrombocytosis and iron deficiency, among others.

VI. Uses of Partial LSP-1, PA-I, and TAP-1 Sequences

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (a) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (b) identify an individual from a minute biological sample (tissue typing); and (c) aid in forensic identification of a biological sample. These applications are described in the subsections below.

a. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the LSP-1, PA-I, and TAP-1, sequences, described herein, can be used to map the location of the LSP-1, PA-I, and TAP-1 gene, respectively, on a chromosome. The mapping of the LSP-1, PA-I, and TAP-1 sequence to chromosomes is an important first step in correlating these sequence with genes associated with disease.

Briefly, the LSP-1, PA-I, and TAP-1 gene can be mapped to a chromosome by preparing PCR primers (preferably 15-25 bp in length) from the LSP-1, PA-I, and TAP-1 sequence. Computer analysis of the LSP-1, PA-I, and TAP-1, sequence can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the LSP-1, PA-I, and TAP-1 sequence will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human

chromosomes. (D'Eustachio P. et al. (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the LSP-1, PA-I, and TAP-1 sequence to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a LSP-1, PA-I, and TAP-1 sequence to its chromosome include *in situ* hybridization (described in Fan, Y. et al. (1990) *PNAS*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma et al., Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York, 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data (such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can

then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. et al. (1987) *Nature,* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the LSP-1, PA-I, and TAP-1 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

b. Tissue Typing

The LSP-1, PA-I, and TAP-1 sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the LSP-1, PA-I, and TAP-1 sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The LSP-1, PA-I, and TAP-1 sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of

about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1, 4, and 7, can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3, 6, and 9, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from LSP-1, PA-I, and TAP-1 sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

c. Use of Partial LSP-1, PA-I, and TAP-1 Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As described above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1, 4, and 7 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the LSP-1, PA-I, and TAP-1 sequences or portions thereof, e.g., fragments derived from the noncoding regions of SEQ ID NO:1, 4, and 7, having a length of at least 20 bases, preferably at least 30 bases.

The LSP-1, PA-I, and TAP-1 sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such LSP-1, PA-I, and TAP-1 probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, e.g., LSP-1, PA-I, and TAP-1 primers or probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application (including the figures) are hereby incorporated by reference.

EXAMPLES

Example 1: Isolation And Characterization of Human LSP-1 cDNAs

In this example, the isolation and characterization of the gene encoding human LSP-1 (also referred to as "HOMEDEPO" or "TANGO 111") is described.

Isolation of the human LSP-1 cDNA

The following methodology takes advantage of the fact that molecules such as LSP-1 have an amino terminal signal sequence which directs certain secreted and membrane-bound proteins through the cellular secretory apparatus.

A partial LSP-1 mRNA was identified by screening of a human bone marrow cDNA library. This library was prepared using mRNA purchased from Clontech, Palo Alto (Cat. no, 6573-1). A signal trap cDNA library was prepared by ligating random primed double stranded cDNA into the expression vector, ptrAP1, resulting in fusions of cDNAs to the reporter, alkaline phosphatase (AP). DNAs from individual clones from this library were prepared by standard techniques and transfected in to human embryonic kidney fibroblasts (293T cells). After 28 hours cell supernatants were collected and assayed for AP activity.

Clones giving rise to detectable AP activity in the supernatants of transfected cells were analyzed further by DNA sequencing and the novel clones subjected to further DNA sequencing.

Once such clone, named LSP-1, was identified. The initial LSP-1 clone contained an open reading frame of 169 amino acids, (fused in-frame to the AP reporter). Blast searching of Genbank with this sequence indicated partial homology to certain immunoglobulin type proteins (illustrated more clearly in the sequence analysis of the full length clone below). A GenBank™ search of the dbEST database utilizing the sequence of this cDNA revealed several EST sequences with greater than 95% nucleotide identity to the partial cDNA.

The LSP-1 nucleic acid molecule was aligned with the FDF03 molecule (described in WO/24906) using the GAP program in the GCG software package (pam120 matrix) and a gap weight of 12 and a length weight of 4. The results showed a 45.3% identity between the two sequences (see Figure 9).

The LSP-1 protein was aligned with the FDF03 protein (described in WO/24906) using the GAP program in the GCG software package (pam120 matrix) and a gap weight of 12 and a length weight of 4. The results showed a 51.6% identity between the two protein sequences (see Figure 10).

Sequencing of Full Length LSP-1 cDNAs

The interesting chromosomal localization (see Example 3) of the LSP-1 genes motivated a search for clones encoding a full length LSP-1 cDNA. Searching of the Genbank database identified several EST sequences that showed a high degree of identity to the partial LSP-1 cDNA identified by signal sequence trapping.

To obtain further sequence data, three ESTs present in the IMAGE clone collection were located and subjected to further DNA sequencing. A single sequence was assembled which completed the LSP-1 open reading frame and extended a 3' UTR and poly A tail. Figure 2 shows the selected clones and details the relationship between the first LSP-1 clone identified by signal trapping, the IMAGE clones and the final composite sequence.

Structure of the LSP-1 protein

The domain structure of the full length LSP-1 proteins is depicted in Figure 3. LSP-1 contains an N-terminal signal peptide (predicted by the signal Pa algorithm), an immunoglobulin-type domain, a transmembrane domain (predicted by MEMSAT, Jones, D.T., Taylor, W.R., and Thornton, J.M. 1994 Biochemistry 33 3038-3049) and a short cytoplasmic domain. The predicted Ig domain is incomplete. However, there is precedent for Ig domains of this kind (Barclay et al., The Leucocyte Antigen Factsbook, Academic Press).

Thus, the LSP-1 protein appears to be a Type 1 membrane protein composed of a single extracellular immunoglobulin type domain, a transmembrane domain and a short cytoplasmic tail. This overall topological structure appears similar to that of other leukocyte membrane proteins, notable CD1, CD3 and CD28 (Barclay et al., The Leucocyte Antigen Factsbook, Academic Press). Also of note is the presence within the predicted LSP-1 membrane spanning domain of a lysine residue. The presence of charged groups within this domain of cell surface proteins suggests that the protein may form homodimers in the cell membrane (Barclay et al., The Leucocyte Antigen Factsbook, Academic Press). Thus, LSP-1 may exist at the cell surface as a homodimer and may exert its effects in this form.

Translation initiation

The translation of the LSP-1 sequence is shown starting from an ATG at nucleotide 1332. This ATG is in a favorable context for translation initiation (good Kozak consensus). Although another in-frame ATG is present upstream at nucleotide 1140, this is not in a favorable context for translation initiation and given the overall predicted topology of the mature protein translated from the second methionine, it is unlikely to be used.

Example 2: Distribution of LSP-1 mRNA In Human Tissues

Probing of northern blots purchased from Clontech revealed a 1.5 kb transcript for LSP-1 only in peripheral blood leukocytes (PBL) and not in any other tissue. Blots contained the following tissues: spleen, thymus, prostate, testis, ovary, small intestine, colon, peripheral blood leukocyte (human multiple tissue northern III, cat. no. 7767-4) spleen, lymph node, thymus, peripheral blood Leucocyte bone marrow and fetal liver (human immune system II cat. no. 7768-1). Also, a larger 4.5 kb transcript was detectable in tissues of the endocrine system (pancreas, adrenal medulla, thyroid, adrenal cortex, testis, thymus, small intestine and stomach-human endocrine blot, cat. no. 7751-1).

PCR analysis of cDNA libraries from various sources was also performed using primers and conditioned as described for chromosomal mapping of LSP-1. PCR detects lower levels of mRNA expression than northern blotting. LSP-1 mRNA was detectable in the following human cDNA libraries; placenta, fetal brain, fetal heart, fetal liver, adult heart, human umbilical vein endothelial cells, HeLa cells, fetal kidney, adult adipose tissue, adult prostate, colorectal adenocarcinoma. lymphocytes, adult lung, adult spleen, HL-60 cells (promyelocytic leukemia), human microvascular endothelial cells and fetal

spleen. LSP-1 mRNA was not detectable in prostrate epithelial cells, megakaryocytes, bronchial epithelial cells or primary osteoblasts.

Example 3: Chromosomal Mapping Of Human LSP-1 Gene

This Example describes the chromosomal mapping of the human and murine LSP-1 genes.

The LSP-1 gene maps to chromosome 7q21-q22, at 111-112 cM (using the Généthon linkage map as reference).

The clone was mapped to two difference Radiation Hybrid (RH) panels, the Stanford Human Genome Center G3 panel and the Genebridge G4 panel, using the following primers (forward: TCACTCAACCAAAACACC (SEQ ID NO:11); reverse CCAGTTCAGAAAGACC (SEQ ID NO:12)). LSP-1 was found to be linked to Genebridge G4 framework marker D7S651, at a distance of 1.7 cR (3000) and a lod score >3; and to Stanford G3 framework marker WI-7004, at a distance of 0 cR (10000) (meaning without recombination between the clone and the marker) with a lod score of 1000. Both Markers are also part of an integrated gene map (http://www.nci.bi.nlm.nih.gov/SCIENCE96/) that serve as a general reference.

The cytogenetic location for LSP-1 was inferred from the map position of close-by genes (Epo and Cytochrome P450 IIIA).

The results of the RH panel mapping place the LSP-1 gene very close to the Erythropoietin (EPO) gene (precise distance unclear due to the lack of resolution of RH mapping). Possibly within 50-100 Kb.

Example 4: Isolation And Characterization Of Murine PA-I cDNAs

In this example, the isolation and characterization of the gene encoding murine Proliferin analog I (PA-I) is described.

The murine gene was discovered by analysis of an EST database (a GenBank™ search of the dbEST database) using human growth hormone as a probe. dbEST clone aa014234 was identified and subsequently obtained from Research Genetics (Huntsville, AL). This EST represents a mouse placenta derived clone which contains an ATG translation initiation codon and is annotated as mouse proliferin-related protein. BlastP searching (BLAST™ searching utilizing an amino acid sequence against a protein database), using the translation product (frame 1) of this sequence, revealed homology to proteins belonging to the prolactin-growth hormone superfamily. The mouse clone was fully sequenced (SEQ ID NO:4).

Example 5: Distribution of PA- I mRNA In Mouse And Human Tissues

Clone aa014234 was digested with a NotI and an EcoRI enzyme to excise the fragment from the vector and this fragment was used as a probe for Northern blots. The fragment was labeled using the Prime It kit from Stratagene (La Jolla, CA) and then hybridized to multi-tissue northern blots from Clontech (Palo Alto, CA) as recommended by the manufacturer. Three blots were used: a human (7760-1), a human immune system II (7768-1) and a mouse embryo (7763-1). A band of about 1Kb was detected in tissue from a mouse day 7 embryo (see Figure 6B) under low stringency hybridization conditions (hybridization at 68 ° C, wash in 2X SSC, 0.05% SDS at 68 ° C for 20 minutes). A band of the same size (1 Kb) was detected in human placenta tissue (see Figure 6A) and in human fetal liver (see Figure 6C) under similar low stringency hybridization conditions (hybridization at 50 ° C, wash in 2X SSC, 0.05% SDS at room temperature for 20 minutes and at 50 ° C for 20 more minutes).

Example 6: Screening for the human cDNA and genomic DNA

A human placenta library (Clonetech), as well as a human fetal liver library (Clonetech) and a human genomic library (Stratagene) are screened using the same probe that was used in the northern blot experiment described above. Hybridization is performed under low stringency conditions. Briefly, the hybridization is performed overnight at 45 ° C, in Church buffer (7% SDS, 250 mM NaHPO4, 2 μM EDTA), and is followed by washing of the filters in 2X SSC, 1% SDS. The blots are exposed to film at -80 ° C for 5 hours. Positive clones are isolated and sequenced using art known techniques.

Example 7: Isolation And Characterization Of Human TAP-1 cDNAs

In this example, the isolation and characterization of the genes encoding human TAP-1 is described.

Contruction of Libraries

Poly A+ RNA from human prostate tissue was used to construct a cDNA library. The cDNA library was constructed by first and second strand synthesis as recommended by the manufacturer for the Gibco BRL kit SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning (Gibco/BRL; Bethesda MD).

Isolation of a human TAP-1 cDNA

A partial human TAP-1 cDNA, also referred to as TANGO-94, was identified by analysis of an EST database using mouse TPO sequence as a probe. A partial human clone (jthqb070d08) was obtained from a human prostate cDNA library and was subsequently fully sequenced. Clone jthqb070d08 was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on October 2, 1997 and has ATCC Accession Number 98554.

A nucleotide sequence of the isolated C-terminal domain of human TAP-1 cDNA (nucleotides 1-523 corresponding to the C-terminus and 3' untranslated sequence) and the predicted amino acid sequence of the human TAP-1 protein (amino acids 1-86) are shown in Figure 7 and in SEQ ID NOs:7 and 8, respectively. The amino acid sequences showed 32% identitity to the C-terminal part of human TPO. The nucleotide sequence corresponding to the coding region of the human TAP-1 cDNA are nucleotides 1-258 of SEQ ID NO:7, nucleotides 259-528 correspond to the 3' untranslated region of the gene.

Isolation of Additional TAP-1 cDNAs

Using jthqb070d08 cDNA as a probe, additional TAP-1 clones were isolated using standard protocols. In brief, clone (jthqb070d08 cDNA) was excised from a pMET vector using SalI and NotI restriction enzymes. The excised fragment was labeled using the Prime It kit from Stratagene (La Jolla, CA) and then hybridized under high stringency conditions to a human fetal liver library. For high stringency conditions, hybridizations were carried out overnight at 65° C in Church buffer. The filters were washed the next day with 2 X SCC / 0.1% SDS. Eight clones were isolated from human fetal liver library and submitted for sequencing. 3 out of the 8 clones contain an insert of approximately 3 kb.

Example 8: Distribution of TAP-1 mRNA In Human Tissues

Northern blots using clone (jthqb070d08 cDNA) were perfomed using standard protocols. In brief, clone (jthqb070d08 cDNA) was excised from a pMET vector using SalI and NotI restriction enzymes. The excised fragment was labeled using the Prime It kit from Stratagene (La Jolla, CA) and then hybridized to multi-tissue northern blots from Clontech (Palo Alto, CA) as recommended by the manufacturer. A strong band was detected at approxiamtely 3 kb in human fetal liver tissues. Additional bands were detected which may be indicative of alternate spliced variants. Two other less intense bands of approximately 5 and 2 kb were detected in all tissues tested.

Example 9: Analysis of TAP-1 amino acid sequence

The partial human TAP-1 cDNA, which is approximately 523 nucleotides in length, and which is approximately 86 amino acid residues in length. The human TAP-1 protein contains four serine-proline-threonine-rich domains. A TAP-1 serine-proline-threonine-rich domain can be found at least, for example, from about amino acids 1-20 of SEQ ID NO:8 (Gly1 to Gly20 of SEQ ID NO:8); from about amino acids 21-40 of SEQ ID NO:8 (Ile20 to Ala40 of SEQ ID NO:8); from about amino acids 41-60 of SEQ ID NO:8 (Val40 to Gly60 of SEQ ID NO:8); and from about amino acids 61-81 of SEQ ID NO:8 (Pro61 to Thr81 of SEQ ID NO:8). The human TAP-1 C-terminal domain appears to encode a secreted protein, e.g., growth factor a secreted protein which shares significant homology, about 32% identity, with the C-terminal region of human TPO.

An alignment of the human TAP-1 amino acid sequences to human TPO sequences is presented in Figure 8. The figure depicts an alignment of the amino acid sequences of TAP-1 (corresponding to amino acids 15 to 75 of SEQ ID NO:8) and human TPO sequences (Swiss-Prot™ Accession Numbers P40225, 1401246, 939627). Identical residues are indicated in the row between the TAP-1 and the TPO sequences by a single amino acid code; conserved amino acid residues are indicated as (+).

Example 10: Expression of LSP-1, PA-I, and TAP-1 in Bacterial Cells

In this example, LSP-1, PA-I, and TAP-1 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, LSP-1, PA-I, and TAP-1 is fused to GST and this fusion polypeptide is expressed in *E. coli*, e.g., strain PEB199. Expression of the GST-LSP-1, PA-I, and TAP-1 fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

Example 11: Expression of Recombinant LSP-1, PA-I, and TAP-1 Protein in COS Cells

To express the LSP-1, PA-I, and TAP-1 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire LSP-1, PA-I, and TAP-1 protein and an HA tag (Wilson et al. (1984) *Cell* 37:767) or a FLAG tag fused in-frame

to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

To construct the plasmid, the LSP-1, PA-I, and TAP-1 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the LSP-1, PA-I, and TAP-1 coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the LSP-1, PA-I, and TAP-1 coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the LSP-1, PA-I, and TAP-1 gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5a, SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

COS cells are subsequently transfected with the LSP-1, PA-I, and TAP-1-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the LSP-1, PA-I, and TAP-1 polypeptide is detected by radiolabelling (^{35}S-methionine or ^{35}S-cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labelled for 8 hours with ^{35}S-methionine (or ^{35}S-cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the LSP-1, PA-I, and TAP-1 coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner

described above, and the expression of the LSP-1, PA-I, and TAP-1 polypeptide is detected by radiolabelling and immunoprecipitation using a LSP-1, PA-I, and TAP-1 specific monoclonal antibody.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed:

1. An isolated nucleic acid molecule selected from the group consisting of:

(a) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:1;

(b) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:3;

(c) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:4;

(d) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:6;

(e) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:7; and

(f) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:9.

2. An isolated nucleic acid molecule which encodes a polypeptide selected from the group consisting of:

(a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2;

(b) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 5; and

(c) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 8.

3. An isolated nucleic acid molecule comprising the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number 98554 or _____.

4. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide selected from the group consisting of:

(a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2;

(b) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 5; and

(c) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 8.

5. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% homologous to the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 601 nucleotides of a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9, or a complement thereof;

c) a nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 60% homologous to the amino acid sequence of SEQ ID NO: 2, 5, or 8; and

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 5, or 8, wherein the fragment comprises at least 15 contiguous amino acid residues of the amino acid sequence of SEQ ID NO: 2, 5, or 8.

6. An isolated nucleic acid molecule which hybridizes to the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 under stringent conditions.

7. An isolated nucleic acid molecule comprising a nucleotide sequence which is complementary to the nucleotide sequence of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.

8. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5, and a nucleotide sequence encoding a heterologous polypeptide.

9. A vector comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.

10. The vector of claim 9, which is an expression vector.

11. A host cell transfected with the expression vector of claim 10.

12. A method of expresing a polypeptide comprising culturing the host cell of claim 11 in an appropriate culture medium to, thereby, express the polypeptide.

13. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 5, or 8, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO: 2, 5, or 8;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 5, or 8, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, 3, 4, 6, 7, or 9 under stringent conditions;

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 60 % homologous to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or 9;

d) a polypeptide comprising an amino acid sequence which is at least 60% homologous to the amino acid sequence of SEQ ID NO:2, 5, or 8.

14. The isolated polypeptide of claim 13 comprising the amino acid sequence of SEQ ID NO: 2, 5, or 8.

15. The polypeptide of claim 13, further comprising heterologous amino acid sequences.

16. An antibody which selectively binds to a polypeptide of claim 13.

17. A method for detecting the presence of a polypeptide of claim 13 in a sample comprising:

a) contacting the sample with a compound which selectively binds to the polypeptide; and

b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of claim 13 in the sample.

18. The method of claim 17, wherein the compound which binds to the polypeptide is an antibody.

19. A kit comprising a compound which selectively binds to a polypeptide of claim 13 and instructions for use.

20. A method for detecting the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in a sample comprising:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample to thereby detect the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in the sample.

21. The method of claim 20, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

22. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 and instructions for use.

23. A method for identifying a compound which binds to a polypeptide of claim 13 comprising:

a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound; and

b) determining whether the polypeptide binds to the test compound.

24. The method of claim 23, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detection of test compound/polypeptide binding;

b) detection of binding using a competition binding assay; and

c) detection of binding using an assay for LSP-1, PA-I, or TAP-1 activity.

25. A method for modulating the activity of a polypeptide of claim 13 comprising contacting the polypeptide or a cell expressing the polypeptide with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

26. A method for identifying a compound which modulates the activity of a polypeptide of claim 13 comprising:

a) contacting a polypeptide of claim 13 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

```
GAATTCGGCACGAGCTGGGGCTCCCCTATGCCTGAGTTCCTGCAACTCAGAAGTTGAGGCCCAGAGATCCAGGTAGAGG   75

GGGCTCCAGTCCTGGGACTGCCCTGCAAGTCATGTGGCCTGGGACAGGGCTTCTCCAAGCTCTGTCCCCTCCTGTGTAG  158

GGCAGGGAGGTCAGAAGGGACCCTGCAGATCCAGTCTCATGTCTGGGGGTCAAGGGGTGGCCTCGAGAGGGACCAGTCT  237

CTGTGTAGGGGACCGTCAGCCCCCTCACCCCTTGAGCAAGACTGTGGTCCCTGCACCAAGGGAGCAGGCCTGGGGTGGG  316

AAGAGGCCAGCTGGGCTGTGGTGGTGCCTGGGGACTGCATGGGAGCCCCTGCCAGGGAGGGAGAGGGACAGAGGACAAC  395

CTGGGGGCTCTGGTGCTTGGGCTGGGGGCTGAGCGCCTGTGACCTCCACTGGCTTCCTCCTTCTCCTCTCTGAGGACTG  474

AATCTGGGGTGCAGCAGAGCACAGACTCAGGCCCGCCTTTCCTCTCCCTGAAAGAGCCTGCGCTGGCCTTGGACAGAGA  553

AAGGGATGAGAAGTGAGGCTGAGTGCGGTGGGGTCTGCAGGGATCCAGGTGGGAGGGGCCCAGCCAGCCAAGGTGAGGC  632

CCAGCCCCTCAGCAGGAAGATGGGCACTGGGGCCCTTGGGCAGGGCTGACTTGACACTTTTGTGTGACTTGGAGCCACT  711

GTGCCCAGCCTGAACACCCTTTCCTGGTAAAACACTCCACAAACCAGGAAGAGAAGGAATATACTGCAACAAAATAAAG  790

GCCAGTCATGCAAGGCCCATGGCTGAAAGTCTTTCAGTCATTTTAGATGAAAGACTGAAATCTTTGCCTCCAAGATCAG  869

GAACAAGAGAAGGATGCCCGCTCTCACTACTTCTATTCAACACAGGATTTGAAGTCAGGCCGGGCACAGTGGCTCACGC  948

CTGTAATCCCAGCACTTTTGGAGGCTGAGGCGGGCAGATTACTTGAGCCTATGAGTGTGAGACCACCCTGGCCAACATG 1027

GCAAAACCCCATCTCTACTAAAAAAAAAAAAAAAAAAAAGGATTTGAAGTCCTGGCCGGAGCAATTAGGCAAGGGATAAAA 1106

AGGCACTAAGGCCCTTTTGCAATAAGAAGCCAGATGGATAAAGGAAGTGCTGGTCACCCTGGAGGTGTACTGGTTTGGG 1185

GAAGGTCCCCGGCCCCCACAGCCCTCTGGGGAGCCTCACCCTGGCTCTCCCCACTCACCTCAGCCCTCAGGCAGCCCCT 1264

                                                                M   G   R     3
CCACAGGACCCCTCTCCTGCCTGGACAGCTCTGCTGGTCTCCCCGTCCCCTGGAGAAGAACAAGGCC ATG GGT CGG 1340

 P   L   L   L   P   L   L   L   L   L   Q   P   P   A   F   L   Q   P   G   G    23
CCC CTG CTG CTG CCC CTG CTG CTC CTG CTG CAG CCG CCA GCA TTT CTG CAG CCT GGT GGC 1400

 S   T   G   S   G   P   S   Y   L   Y   G   V   T   Q   P   K   H   L   S   A    43
TCC ACA GGA TCT GGT CCA AGC TAC CTT TAT GGG GTC ACT CAA CCA AAA CAC CTC TCA GCC 1460

 S   M   G   G   S   V   E   I   P   F   S   F   Y   Y   P   W   E   L   A   T    63
TCC ATG GGT GGC TCT GTG GAA ATC CCC TTC TCC TTC TAT TAC CCC TGG GAG TTA GCC ACA 1520

 A   P   D   V   R   I   S   W   R   R   G   H   F   H   G   Q   S   F   Y   S    83
GCT CCC GAC GTG AGA ATA TCC TGG AGA CGG GGC CAC TTC CAC GGG CAG TCC TTC TAC AGC 1580

 T   R   P   P   S   I   H   K   D   Y   V   N   R   L   F   L   N   W   T   E   103
ACA AGG CCG CCT TCC ATT CAC AAG GAT TAT GTG AAC CGG CTC TTT CTG AAC TGG ACA GAG 1640

 G   Q   E   S   G   F   L   P   I   S   N   L   R   K   E   D   Q   S   V   V   123
GGT CAG GAG AGC GGC TTC CTC AGG ATC TCA AAC CTG CGG AAG GAG GAC CAG TCT GTG TAT 1700

 F   C   R   V   E   L   D   T   R   R   S   G   R   Q   Q   L   Q   S   I   K   143
TTC TGC CGA GTC GAG CTG GAC ACC CGG AGA TCA GGG AGG CAG CAG TTG CAG TCC ATC AAG 1760

 G   T   K   L   T   I   T   Q   A   V   T   T   T   T   T   W   R   P   S   S   163
```

FIGURE 1

```
GGG ACC AAA CTC ACC ATC ACC CAG GCT GTC ACA ACC ACC ACC ACC TGG AGG CCC AGC AGC  1820

 T   T   T   I   A   G   L   R   V   T   E   S   K   G   H   S   E   S   W   H   183
ACA ACC ACC ATA GCC GGC CTC AGG GTC ACA GAA AGC AAA GGG CAC TCA GAA TCA TGG CAC  1880

 L   S   L   D   T   A   I   R   V   A   L   A   V   A   V   L   K   T   V   I   203
CTA AGT CTG GAC ACT GCC ATC AGG GTT GCA TTG GCT GTC GCT GTG CTC AAA ACT GTC ATT  1940

 L   G   L   L   C   L   L   L   W   W   R   R   R   K   G   S   R   A   P   S   223
TTG GGA CTG CTG TGC CTC CTC CTG TGG TGG AGG AGA AGG AAA GGT AGC AGG GCG CCA AGC  2000

 S   D   F   *
AGT GAC TTC TGA                                                                 227
                                                                                2012

CCAACAGAGTGTGGGGAGAAGGGATGTGTATTAGCCCCGGAGGACGTGATGTGAGACCCGCTTGTGAGTCCTCCACACT  2091

CGTTCCCCATTGGCAAGATACATGGAGAGCACCCTGAGGACCTTTAAAAGGCAAAGCCGCAAGGCAGAAGGAGGCTGGG  2170

TCCCTGAATCACCGACTGGAGGAGAGTTACCTACAAGAGCCTTCATCCAGGAACATCCACACTGCAATGATATAGGAAT  2249

GAAGTCTGAACTCCACTGAATTAAACCACTGGCATTTGGGGGCTGTTCATTATAGCAGTGCAAAGAGTTCCTTTATCCT  2328

CCCCAAGGATGGAAAATACAATTTATTTTGCTTACCATACACCCCTTTTCTCCTCGTCCACATTTTCCAATCTGTATGG  2407

TGGCTGTCTTCTATGGCAAAAGTTTTGGGGAATAAATAACGTTAAATGCTGCTGA                         2462
```

FIGURE 1 (cont'd)

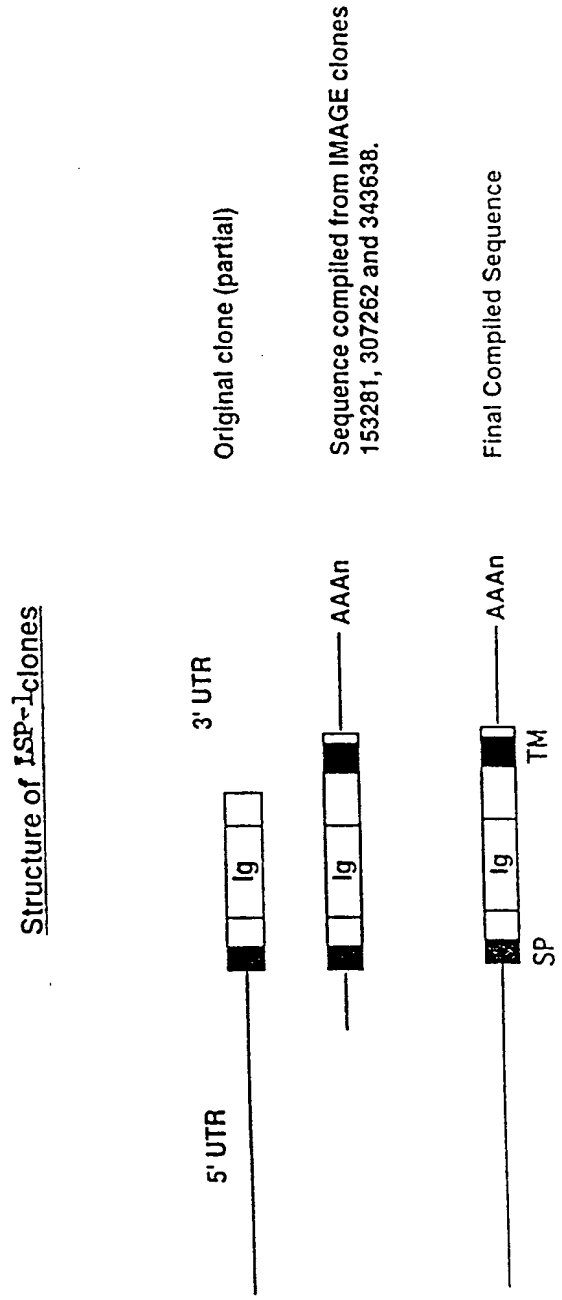
Structure of ISP-1 clones
3' UTR
Original clone (partial)
Sequence compiled from IMAGE clones
153281, 307262 and 343638.
Final Compiled Sequence
Ig
AAAn
5' UTR
SP
TM
Genbank Accession Numbers of IMAGE clones; 153281 - R50327 (5')
307262 - W21480 (5'), N93437 (3')
343638 - W69552 (5'), W69468 (3')

FIGURE 2

Domain structure of LSP-1

```
1   MGRPLLLPLL LLLQPPAFLQ PGGSTGSGPS YLYGVTQPKH 40
41  LSASMGGSVE LPFSFYYPWE LATAPDVRLS WRRGHFHGQS 80
81  FYSTRPPRSH HKDYVNRLFL NWTEGQESGF LRLSNLRKED 120
121 SVYFCRVELD TRRSGRQQLQ SIKGTKLTIT QAVTTTTTWR 160
161 PSSTTTIAGL RVTESKGHSE SWHLSLDTAI RVALAVAVLK 200
201 TVILGLLCLL LWWRRRKGSR APSSDF
```

1-20 (underline) = Signal peptide
46 - 128 (shaded box) = Ig domain
ASN-100 (*) = N-linked glycosylation site (NWTE)
192-213 (broken line) = Transmembrane domain
216 - 219 (open box) = cAMP/cGMP kinase phosphorylation site

FIGURE 3

```
                                                                  M  S  F  S  F  S      6
TCGACGCGCCTATTCTTGCACGAGGCAGACAGGCTGTGCCAGAACTCTTCAGAG ATG TCA TTT TCT TTC TCT     72

 Q  P  C  P  S  G  A  L  L  L  V  V  V  S  S  L  L  L  W  E     26
CAA CCA TGC CCC TCA GGG GCA CTT CTG CTG GTG GTG GTG TCA AGC CTC CTT TTA TGG GAG    132

 N  V  A  S  V  P  L  S  S  N  E  T  D  G  Y  P  L  S  I  N     46
AAT GTG GCC TCT GTA CCT TTG AGT AGC AAT GAG ACT GAT GGT TAT CCA TTA TCC ATC AAT    192

 G  L  F  H  N  A  M  R  L  T  W  N  I  K  N  L  N  M  E  L     66
GGG CTG TTT CAT AAT GCC ATG AGA CTA ACT TGG AAT ATC AAA AAC CTC AAC ATG GAA CTG    252

 R  K  T  Y  T  V  N  Q  V  S  E  K  L  Y  E  N  Y  M  L  D     86
CGC AAG ACA TAT ACA GTC AAT CAA GTC TCT GAA AAA TTA TAC GAG AAC TAT ATG CTT GAC    312

 F  I  E  D  M  E  Y  L  V  K  A  L  T  C  C  H  N  Y  S  I    106
TTT ATT GAG GAC ATG GAG TAT CTG GTC AAG GCT CTC ACC TGC TGC CAC AAT TAT TCC ATC    372

 K  T  P  E  N  L  D  E  A  Q  Q  I  P  F  N  E  F  P  K  L    126
AAA ACT CCA GAA AAC CTG GAC GAA GCT CAA CAG ATT CCT TTT AAC GAA TTT CCA AAG CTG    432

 I  L  S  R  M  W  A  W  N  E  T  S  K  V  L  L  T  T  L  R    146
ATC CTC AGT AGA ATG TGG GCT TGG AAT GAA ACT TCT AAA GTT CTA CTG ACC ACA CTC AGA    492

 S  I  P  G  M  H  D  D  V  I  S  L  A  K  N  I  E  T  K  L    166
AGT ATT CCA GGA ATG CAT GAT GAT GTC ATT TCA TTA GCC AAA AAC ATT GAA ACA AAA CTT    552

 A  E  L  F  E  Y  T  Q  S  I  L  N  S  I  Y  G  T  T  T  T    186
GCA GAG CTT TTT GAG TAC ACC CAG AGT ATA CTC AAC TCG ATT TAT GGA ACA ACA ACA ACA    612

 G  N  V  E  Y  T  V  F  S  G  L  E  D  L  K  S  S  D  E  E    206
GGA AAT GTG GAA TAC ACC GTC TTT TCT GGT CTT GAA GAC TTA AAA TCA TCT GAT GAA GAA    672

 F  S  L  F  D  L  C  K  F  S  Y  C  L  R  V  D  I  H  M  V    226
TTT AGT CTT TTT GAC CTT TGT AAA TTT TCC TAT TGC TTA CGT GTA GAT ATA CAT ATG GTT    732

 E  L  Y  L  K  L  L  E  C  V  V  Y  V  S  S  D  V  C  L  S    246
GAA CTT TAT CTC AAG CTA TTA GAG TGT GTG GTA TAT GTT AGT AGT GAT GTT TGT TTA TCC    792

 K  N  I  R  D  A  S  *                                          254
AAA AAT ATT AGA GAT GCT TCA TGA                                   816

TGCTGAATCTTTTTTAAATAATCTTAATTTTTATAATTGTGAAAGTATAATTGAGTATAACGAGTGTCTTTTAAAATAAAA  895

ATAAACTATATATATAAAAAAAAAAAAAAAAAAAAAAAA                           933
```

FIGURE 4

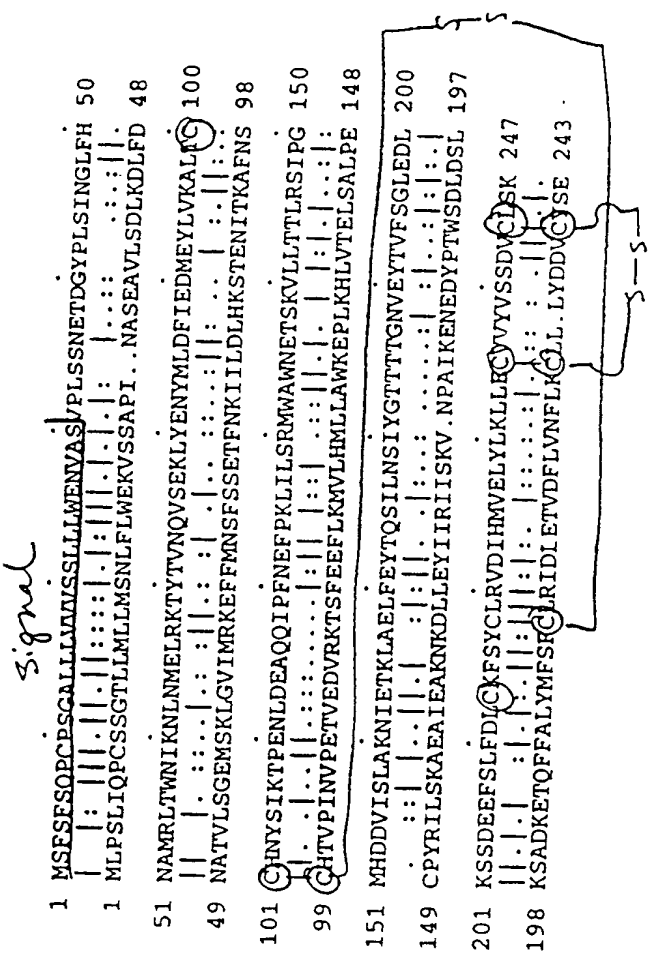

FIGURE 5

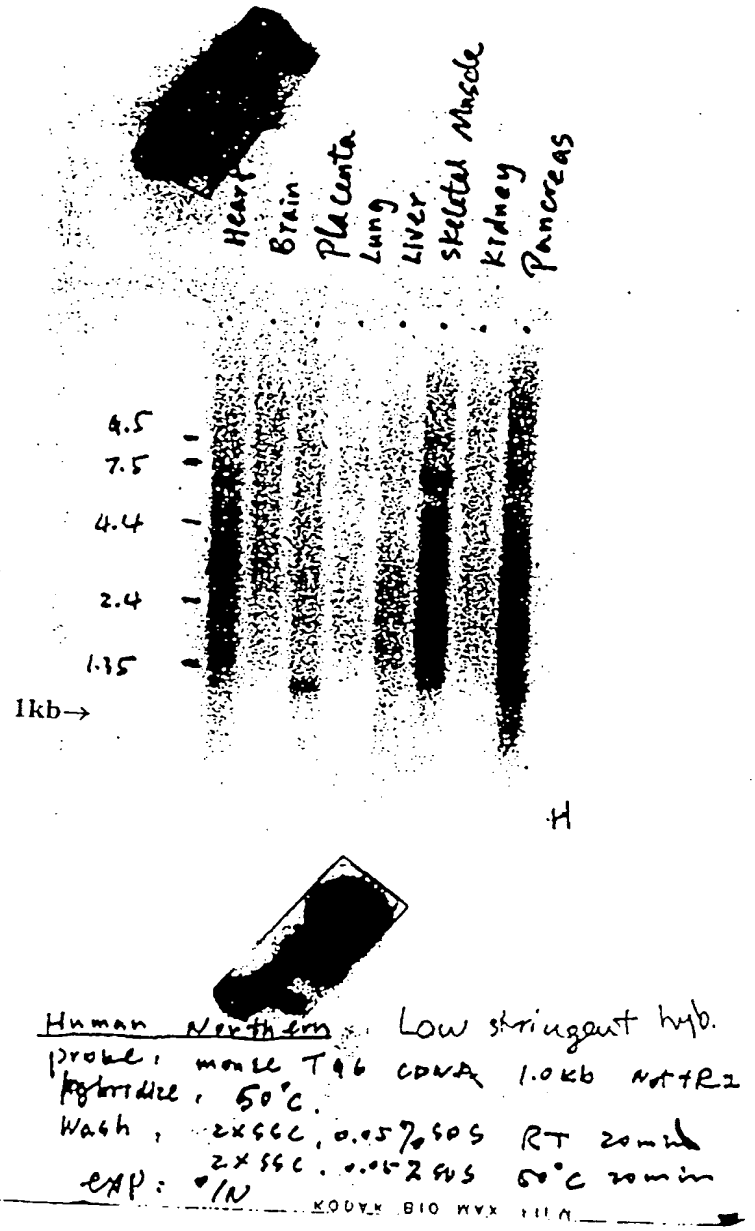

FIGURE 6

q18197
Mouse Embryo Northern
Probe: Mouse TAS cDNA 0.9 kb
hyb: 68°C
Wash: 2XSSC 0.05% SDS 68°C 20min
exp: O/N

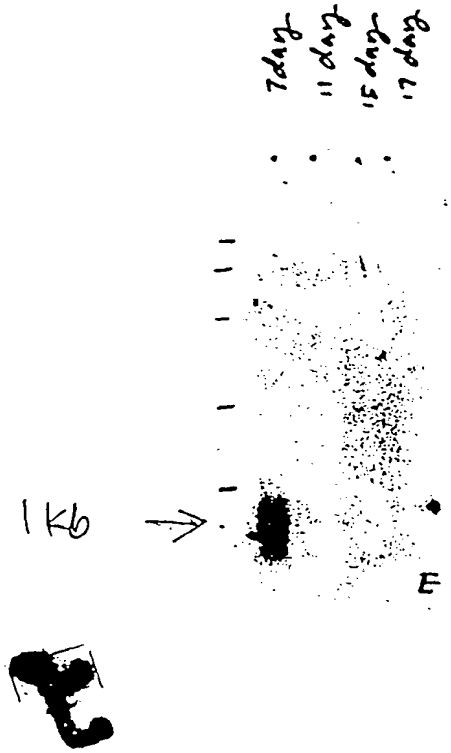

FIGURE 6 (cont'd)

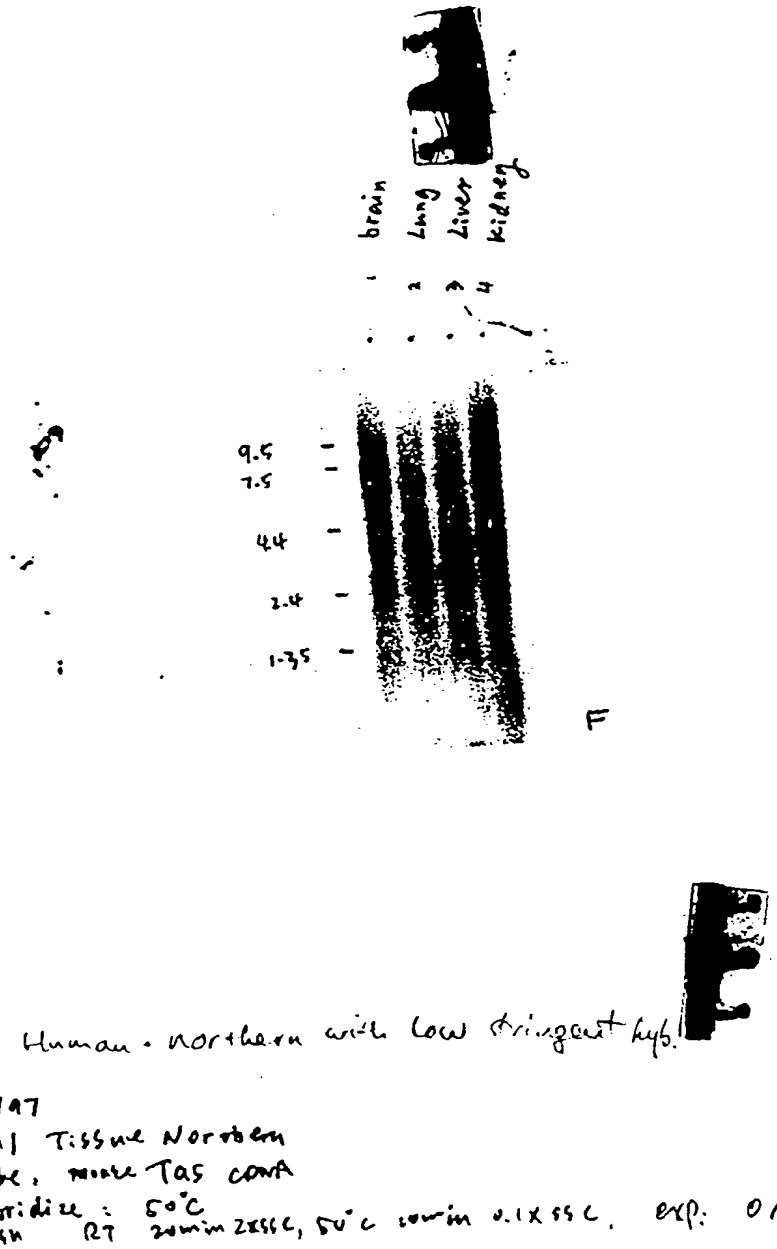

FIGURE 6 (cont'd)

```
 G  S  T  P  A  S  A  V  S  A  R  C  A  T  S  S  R  S  L  G
20
GGA TCG ACC CCC GCG TCC GCG GTT TCT GCG CGG TGC GCG ACC TCG TCC CGA AGC CTG GGG
60

 I  H  P  L  E  S  P  L  S  P  S  V  K  V  E  P  L  T  V  A
40
ATA CAC CCT CTC GAG AGC CCG CTG TCG CCC TCC GTT AAG GTC GAA CCC CTC ACA GTT GCT
120

 V  G  N  S  S  P  T  F  P  R  S  G  S  R  P  I  G  K  L  G
60
GTG GGC AAC TCC AGC CCA ACA TTC CCT CGC TCT GGT TCT CGC CCC ATT GGG AAA CTC GGC
180

 P  T  L  P  T  F  L  D  E  V  S  P  F  S  P  L  K  C  Q  I
80
CCC ACG CTT CCC ACT TTT CTG GAT GAG GTG TCC CCT TTC TCC CCA CTA AAA TGT CAA ATA
240

 T  Y  G  G  S  S  *
87
ACC TAC GGA GGG TCT TCC TGA
261

AACCCGCAGAGGAAAATTCAGGCAACGCCTCGCGTTGTGGTTTCAGGCTGCATGTACCAAGTAGTTCAGACGRTTTGGC
340

TCGGATGGAAAAAAAATCTTCTGCAATTAMTTTCCAATTCCTAAGTCTTCTTGGAAATCTTATACCACTAGTTCAATCTT
419

CAAGTCATGTCTGATGCTTTGAAAGGGRATTACAGGRAAACCAGTTCAAGTTAMTTTTCAGACCAGATTTCCAGCTCTT
498

CCACAAGTGCATCATTTCAATTGCCCATTT
528
```

FIGURE 7

```
HARPO:SSRSLGIHPLESPLSPSVKVEPLTVAVGNSSPTFPRSGSRPIGKLGPTLPTFLDEVSPFSP
      S R+LG   + S  S +  + P      + SPT P +G   +  L PTLPT + ++ P  P
TPO:  SRRTLGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLPPTLPTPVVQLHPLLP
```

FIGURE 8

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989)
> T111 Cmhe246a6cons - Repeat Mask                    1426 bp vs.
> PPN:AC01554|WO98/24906|schering corp|isolated m    1249 bp
scoring matrix: pam120.mat, gap penalties: -12/-4
45.3% identity;          Global alignment score: 1812

                  10            20            30            40
inputs GAAT----TCGGCACGAG----CTGGGGCTCCC--CTATGCCTGAGTTCCTGCA----AC
       :..:    . :::.: .:    : . .:: : :  : ..:::::::  :: ::.    .:
       GTTTGGGGAAGGCTCCTGGCCCCCACAGCCCTCTTCGGAGCCTGAGC-CCGGCTCTCCTC
                10         20         30         40         50

          50         60         70         80         90          100
inputs TCAGAAGTTGAGGCCCAGAGATCCAGGTAGAGGGGGCTCCAGTCCTG---GGACTGCCCT
       .:. :  : .:  :::::. . ::   :  : .:::: ::. :::::   :::: :: ::
       ACTCACCTCAACCCCCAGGCGGCCCC-TCCACAGGGCCCCTCTCCTGCCTGGACGGCTCT
      60         70         80          90        100        110

            110           120        130        140        150        160
inputs GCAAGTCAT---GTGGCCTGGGACAGGGCTTCTCCAAGCTCTGTCCCCTCCTGTGTAGGG
       ::..:::.    ::  ::::::.. ::..:..  :::.:    : ::::: :::
       GCTGGTCTCCCCGTCCCCTGGAGAAGAACAAGGCCATGGGTCGGCCCCTGCTG--------
      120        130        140        150        160        170

               170        180        190        200        210
inputs CAGGGAGGTCAGAAGGGACCCTGCAGATCCAGTCT-CATGTCTGGGGGTCAAGGGGTGGC
       :.     : :  .:  : :::::::.: : : : :. ::: :::: .:  :.:: ::   :
       CT-----GCCCCTACTGCCCCTGCTGCTGCCGCCAGCATTTCTGCAG-CCTAGTGGCT-C
                  180        190        200        210        220

    220          230        240        250        260        270
inputs CTCGAGAG--GGACCAGTCT-CTGTGTAGGGGACCGTCAGCCCCCTCACCCCTTGAGCAA
       :.:..::   ::.:::. :: :  : ..:::::.:  :::.::   .:::: ::
       CACAGGATCTGGTCCAAGCTACCTTTATGGGGTCACTCAACCAAAACACCTCT-------
          230        240        250        260        270

         280        290        300        310        320        330
inputs GACTGTGGTCCCTGCACCAAGGGAGCAGGC-CTGGGGTGGGAAGAGGCCAGCTGGGCTGT
         :.:      :::   :::.:::.:   :: ::: ::    ::.  ::. :.:  ::
       --CAG-----CCT---CCATGGGTG---GCTCTGTGG----AAATCCCCTTCT--CCTTC
         280                290           300           310

          340         350        360        370         380        390
inputs GGTGGTGCCTGGG-GACTGCATGGGAGCCCCTGCCAGGGAGGGA--GAGGGACAGAGG-A
       .: .  :::::: :. .:: . .:  :::  .: .:.::. ..   .:.:::.:.:  :
       TATTACCCCTGGGAGTTAGCCACAGCTCCC-GACGNGAGAATATCCTGGAGACGGGGCCA
      320        330        340        350        360        370

            400        410        420        430        440        450
inputs CAACCTGGGGGCTCTGGTGCTTGGGCTGGGGGCTGAGCGCCTGTGACC-TCCACTGGCTT
       :..::. ::: :. :  : ::. .::. ..::::    :::::   .:: : :::..: .:
       CTTCCACGGG-CAGTCCTTCTACAGCACAAGGC----CGCCT---TCCATTCACAAGGAT
       380        390        400           410          420

              460        470        480        490        500
inputs CCTCCTTCTC--CTCTCTGAGGACTGAATCTGGGGTGCAGCAGAGCA-CAGACTCAGGCC
```

FIGURE 9

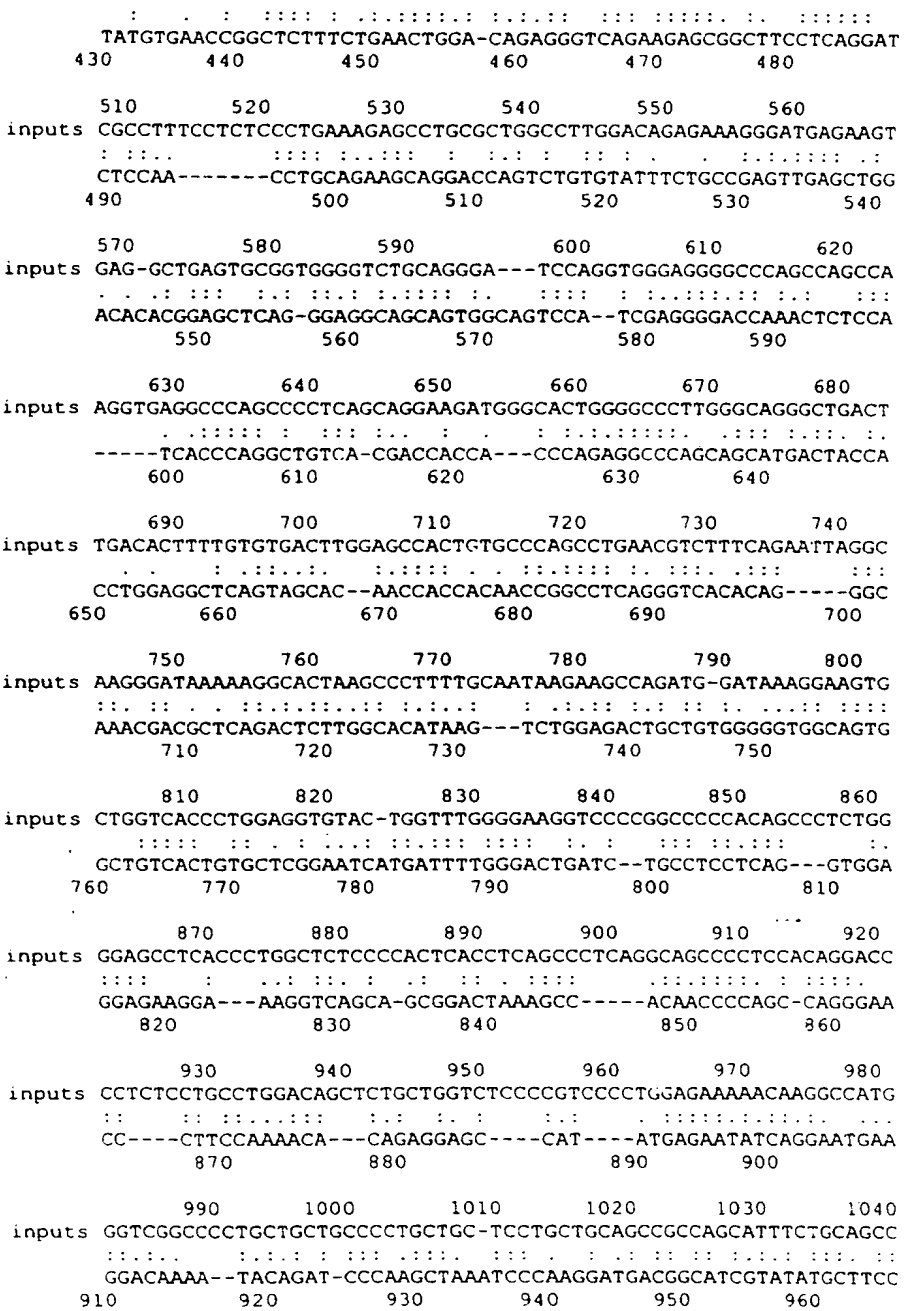

FIGURE 9 (cont'd)

```
              1050       1060       1070       1080       1090       1100
inputs TGGTGGCTCCACAGGATCTGGTCCAAGCTACCTTTATGGGGTCACTCAACCAAAACACCT
          : :: ::.:           :::: ::: ::.          : :::: :::.:.::::
       C--TTGC-CCTC---------TCCA-GCT-CCA---------C-CTCACCCAGAGCACCT
          970                  980                  990       1000

              1110       1120       1130       1140       1150       1160
inputs CTCAGCCTCCATGGGTGGCTCTGTGGAAATCCCCTTCTCCTTCTATTACCCCTGGGAGTT
       : ::::::.:::.                  :::::        :.: ..:::: .:.:  .
       CCCAGCCACCGT------------------CCCCT-------CNAGAGCCCCCAGAACGA
           1010                                1020       1030

              1170       1180       1190       1200       1210
inputs AGCCACAGCTCCC-GACGTGAGAA--TATCCTGGAGACGGGGCCACTTCCACGGGCAGTC
       ..:: :.: .: : :.: :.:...  ::.::.. .::::.:  ::.::  :: .:.:.:.
       GACC-CTGTACTCTGTCTTAAAGGCCTAACCAATGGACAGC-CCTCT--CA-AGACTGAA
      1040       1050       1060       1070       1080       1090

      1220       1230       1240       1250       1260       1270
inputs CTTCTACAGCACAAGGCCGCCTTCCATTCACAAGGATTATGTGAACCGGCTCTTTCTGAA
        :  :. .:: :::.:  :.  :  :. .::: .: :.:        ::.::: . ::::::
       -TGGTGAGGC-CAGGTACAG-TGGCGCACACCTGTAATC------CCAGCT-ACTCTGAA
            1100       1110       1120             1130       1140

      1280       1290       1300       1310       1320       1330
inputs CTGGACAGAGGGTCAGGAGAGCGGCTTCCTCAGGATCTCAAACCTGCGGAAGGAGGACCA
          : :.:::: ::::.:           :::.:.   :   ::.: :...  :::.:::
       ---GCCTGAGG--CAGAA-----------TCAAGTGAGC---CCAG-GAGTTCAGGGCCA
             1150                   1160          1170       1180

      1340       1350       1360       1370       1380       1390
inputs GTCTGTGTATTTCTGCCGAGTCGAGCTGGACACCCGGAGATCAGGGAGGCAGCAGTTGCA
       : :: ::.....:     :.. :::: ::    ::   :  ::.: ... :. : .:. .
       G-CTTTGATAAT-----GGAGCGAGATG----CC---ATCTCTAGTTAAAAATATATATT
            1190           1200             1210       1220

      1400       1410       1420
inputs GTCCATCAAGGGGACCAAACGGCCTCGAG
       ..: :: ::.: .:: ::.         .
       AACAAT-AAAGTAACAAAT--------TT
    1230       1240
```

FIGURE 9 (cont'd)

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989)
> Tangol11aa                                              171 aa vs.
> AC01554aa                                               304 aa
scoring matrix: pam120.mat, gap penalties: -12/-4
51.6% identity;          Global alignment score: 291

                10        20        30        40        50        60
inputs MGRPLLLPLLLLLQPPAFLQPGGSTGSGPSYLYGVTQPKHLSASMGGSVEIPFSFYYPWE
       :::::::::: :: :::::::.:::::::::::::::::::::::::::::::::::::
       MGRPLLLPLLPLLLPPAFLQPSGSTGSGPSYLYGVTQPKHLSASMGGSVEIPFSFYYPWE
                10        20        30        40        50        60

                70        80        90       100       110       120
inputs LATAPDVRISWRRGHFHGQSFYSTRPPSIHKDYVNRLFLNWTEGQESGFLRISNLRKEDQ
       ::::::.:::::::::::::::::::::::::::::::::: ::::::::::.:.::.
       LATAPDXRISWRRGHFHGQSFYSTRPPSIHKDYVNRLFLNWTEGQKSGFLRISNLQKQDQ
                70        80        90       100       110       120

               130       140       150                 160       170
inputs SVYFCRVELDTRRSGRQQLQSIKGTKLTITQAVTTTT--------TWRPSSTTTITGL--
       :::::::::::: ::::: ::: :::::.:::::::::         ::: ::::::.:::
       SVYFCRVELDTRSSGRQQWQSIEGTKLSITQAVTTTTQRPSSMTTTWRLSSTTTTTGLRV
               130       140       150       160       170       180


inputs ------------------------------------------------------------

       TQGKRRSDSWHISLETAVGVAVAVTVLGIMILGLICLLRWRRRKGQQRTKATTPAREPFQ
               190       200       210       220       230       240


inputs ------------------------------------------------------------

       NTEEPYENIRNEGQNTDPKLNPKDDGIVYASLALSSSTSPRAPPSHRPLXSPQNETLYSV
               250       260       270       280       290       300


inputs ---E

       LKAN
```

FIGURE 10

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: MILLENNIUM BIOTHERAPEUTICS, INC.
(B) STREET: 640 MEMORIAL DRIVE
(C) CITY: CAMBRIDGE
(D) STATE: MASSACHUSETTS
(E) COUNTRY: US
(F) POSTAL CODE (ZIP): 02139
(G) TELEPHONE:
(H) TELEFAX:

(ii) TITLE OF INVENTION: SIGNAL PEPTIDE CONTAINING PROTEINS AND USES THEREFOR

(iii) NUMBER OF SEQUENCES: 12

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: LAHIVE & COCKFIELD, LLP
(B) STREET: 28 STATE STREET
(C) CITY: BOSTON
(D) STATE: MASSACHUSETTS
(E) COUNTRY: US
(F) ZIP: 02109

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US98/
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 60/061,143
(B) FILING DATE: OCTOBER 6, 1997
(C) APPLICATION NUMBER: US 60/061,149
(D) FILING DATE: OCTOBER 6, 1997
(E) APPLICATION NUMBER: 60/061,159
(F) FILING DATE: OCTOBER 6, 1997
(G) APPLICATION NUMBER: 09/004,206
(H) FILING DATE: JANUARY 8, 1998
(I) APPLICATION NUMBER: 09/010,674
(J) FILING DATE: JANUARY 22, 1998
(K) APPLICATION NUMBER: 09/014,347
(L) FILING DATE: JANUARY 27, 1998

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: MANDRAGOURAS, AMY E.
(B) REGISTRATION NUMBER: 36,207
(C) REFERENCE/DOCKET NUMBER: MEI-004CPPC

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617)227-7400
(B) TELEFAX: (617)742-4214

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2462 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1332..2009

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAATTCGGCA CGAGCTGGGG CTCCCCTATG CCTGAGTTCC TGCAACTCAG AAGTTGAGGC 60

CCAGAGATCC AGGTAGAGGG GGCTCCAGTC CTGGGACTGC CCTGCAAGTC ATGTGGCCTG 120

GGACAGGGCT TCTCCAAGCT CTGTCCCCTC CTGTGTAGGG CAGGGAGGTC AGAAGGGACC 180

CTGCAGATCC AGTCTCATGT CTGGGGGTCA AGGGGTGGCC TCGAGAGGGA CCAGTCTCTG 240

TGTAGGGGAC CGTCAGCCCC CTCACCCCTT GAGCAAGACT GTGGTCCCTG CACCAAGGGA 300

GCAGGCCTGG GGTGGGAAGA GGCCAGCTGG GCTGTGGTGG TGCCTGGGGA CTGCATGGGA 360

GCCCCTGCCA GGGAGGGAGA GGGACAGAGG ACAACCTGGG GGCTCTGGTG CTTGGGCTGG 420

GGGCTGAGCG CCTGTGACCT CCACTGGCTT CCTCCTTCTC CTCTCTGAGG ACTGAATCTG 480

GGGTGCAGCA GAGCACAGAC TCAGGCCCGC CTTTCCTCTC CCTGAAAGAG CCTGCGCTGG 540

CCTTGGACAG AGAAAGGGAT GAGAAGTGAG GCTGAGTGCG GTGGGGTCTG CAGGGATCCA 600

GGTGGGAGGG GCCCAGCCAG CCAAGGTGAG GCCCAGCCCC TCAGCAGGAA GATGGGCACT 660

GGGGCCCTTG GGCAGGGCTG ACTTGACACT TTTGTGTGAC TTGGAGCCAC TGTGCCCAGC 720

CTGAACACCC TTTCCTGGTA AAACACTCCA CAAACCAGGA AGAGAAGGAA TATACTGCAA 780

CAAAATAAAG GCCAGTCATG CAAGGCCCAT GGCTGAAAGT CTTTCAGTCA TTTTAGATGA 840

AAGACTGAAA TCTTTGCCTC CAAGATCAGG AACAAGAGAA GGATGCCCGC TCTCACTACT 900

TCTATTCAAC ACAGGATTTG AAGTCAGGCC GGGCACAGTG GCTCACGCCT GTAATCCCAG 960

CACTTTTGGA GGCTGAGGCG GGCAGATTAC TTGAGCCTAT GAGTGTGAGA CCACCCTGGC 1020

```
CAACATGGCA AAACCCCATC TCTACTAAAA AAAAAAAAAA AAAAGGATTT GAAGTCCTGG    1080

CCGGAGCAAT TAGGCAAGGG ATAAAAAGGC ACTAAGGCCC TTTTGCAATA AGAAGCCAGA    1140

TGGATAAAGG AAGTGCTGGT CACCCTGGAG GTGTACTGGT TTGGGGAAGG TCCCCGGCCC    1200

CCACAGCCCT CTGGGGAGCC TCACCCTGGC TCTCCCCACT CACCTCAGCC CTCAGGCAGC    1260

CCCTCCACAG GACCCCTCTC CTGCCTGGAC AGCTCTGCTG GTCTCCCCGT CCCCTGGAGA    1320

AGAACAAGGC C ATG GGT CGG CCC CTG CTG CTG CCC CTG CTG CTC CTG CTG      1370
             Met Gly Arg Pro Leu Leu Leu Pro Leu Leu Leu Leu Leu
               1               5                  10

CAG CCG CCA GCA TTT CTG CAG CCT GGT GGC TCC ACA GGA TCT GGT CCA      1418
Gln Pro Pro Ala Phe Leu Gln Pro Gly Gly Ser Thr Gly Ser Gly Pro
     15                  20                  25

AGC TAC CTT TAT GGG GTC ACT CAA CCA AAA CAC CTC TCA GCC TCC ATG      1466
Ser Tyr Leu Tyr Gly Val Thr Gln Pro Lys His Leu Ser Ala Ser Met
 30                  35                  40                  45

GGT GGC TCT GTG GAA ATC CCC TTC TCC TTC TAT TAC CCC TGG GAG TTA      1514
Gly Gly Ser Val Glu Ile Pro Phe Ser Phe Tyr Tyr Pro Trp Glu Leu
                 50                  55                  60

GCC ACA GCT CCC GAC GTG AGA ATA TCC TGG AGA CGG GGC CAC TTC CAC      1562
Ala Thr Ala Pro Asp Val Arg Ile Ser Trp Arg Arg Gly His Phe His
             65                  70                  75

GGG CAG TCC TTC TAC AGC ACA AGG CCG CCT TCC ATT CAC AAG GAT TAT      1610
Gly Gln Ser Phe Tyr Ser Thr Arg Pro Pro Ser Ile His Lys Asp Tyr
         80                  85                  90

GTG AAC CGG CTC TTT CTG AAC TGG ACA GAG GGT CAG GAG AGC GGC TTC      1658
Val Asn Arg Leu Phe Leu Asn Trp Thr Glu Gly Gln Glu Ser Gly Phe
     95                  100                 105

CTC AGG ATC TCA AAC CTG CGG AAG GAG GAC CAG TCT GTG TAT TTC TGC      1706
Leu Arg Ile Ser Asn Leu Arg Lys Glu Asp Gln Ser Val Tyr Phe Cys
110                 115                 120                 125

CGA GTC GAG CTG GAC ACC CGG AGA TCA GGG AGG CAG CAG TTG CAG TCC      1754
Arg Val Glu Leu Asp Thr Arg Arg Ser Gly Arg Gln Gln Leu Gln Ser
                130                 135                 140

ATC AAG GGG ACC AAA CTC ACC ATC ACC CAG GCT GTC ACA ACC ACC ACC      1802
Ile Lys Gly Thr Lys Leu Thr Ile Thr Gln Ala Val Thr Thr Thr Thr
            145                 150                 155

ACC TGG AGG CCC AGC AGC ACA ACC ACC ATA GCC GGC CTC AGG GTC ACA      1850
Thr Trp Arg Pro Ser Ser Thr Thr Thr Ile Ala Gly Leu Arg Val Thr
        160                 165                 170
```

```
GAA AGC AAA GGG CAC TCA GAA TCA TGG CAC CTA AGT CTG GAC ACT GCC      1898
Glu Ser Lys Gly His Ser Glu Ser Trp His Leu Ser Leu Asp Thr Ala
    175                 180                 185

ATC AGG GTT GCA TTG GCT GTC GCT GTG CTC AAA ACT GTC ATT TTG GGA      1946
Ile Arg Val Ala Leu Ala Val Ala Val Leu Lys Thr Val Ile Leu Gly
190                 195                 200                 205

CTG CTG TGC CTC CTC CTG TGG TGG AGG AGA AGG AAA GGT AGC AGG GCG      1994
Leu Leu Cys Leu Leu Leu Trp Trp Arg Arg Arg Lys Gly Ser Arg Ala
                210                 215                 220

CCA AGC AGT GAC TTC TGACCAACAG AGTGTGGGGA GAAGGGATGT GTATTAGCCC      2049
Pro Ser Ser Asp Phe
            225

CGGAGGACGT GATGTGAGAC CCGCTTGTGA GTCCTCCACA CTCGTTCCCC ATTGGCAAGA    2109

TACATGGAGA GCACCCTGAG GACCTTTAAA AGGCAAAGCC GCAAGGCAGA AGGAGGCTGG    2169

GTCCCTGAAT CACCGACTGG AGGAGAGTTA CCTACAAGAG CCTTCATCCA GGAACATCCA    2229

CACTGCAATG ATATAGGAAT GAAGTCTGAA CTCCACTGAA TTAAACCACT GGCATTTGGG    2289

GGCTGTTCAT TATAGCAGTG CAAAGAGTTC CTTTATCCTC CCCAAGGATG GAAAATACAA    2349

TTTATTTTGC TTACCATACA CCCCTTTTCT CCTCGTCCAC ATTTTCCAAT CTGTATGGTG    2409

GCTGTCTTCT ATGGCAAAAG TTTTGGGGAA TAAATAACGT TAAATGCTGC TGA           2462
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 226 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Gly Arg Pro Leu Leu Leu Pro Leu Leu Leu Leu Leu Gln Pro Pro
 1               5                  10                  15

Ala Phe Leu Gln Pro Gly Gly Ser Thr Gly Ser Gly Pro Ser Tyr Leu
            20                  25                  30

Tyr Gly Val Thr Gln Pro Lys His Leu Ser Ala Ser Met Gly Gly Ser
        35                  40                  45

Val Glu Ile Pro Phe Ser Phe Tyr Tyr Pro Trp Glu Leu Ala Thr Ala
    50                  55                  60
```

```
Pro Asp Val Arg Ile Ser Trp Arg Arg Gly His Phe His Gly Gln Ser
 65                  70                  75                  80

Phe Tyr Ser Thr Arg Pro Pro Ser Ile His Lys Asp Tyr Val Asn Arg
                 85                  90                  95

Leu Phe Leu Asn Trp Thr Glu Gly Gln Glu Ser Gly Phe Leu Arg Ile
            100                 105                 110

Ser Asn Leu Arg Lys Glu Asp Gln Ser Val Tyr Phe Cys Arg Val Glu
        115                 120                 125

Leu Asp Thr Arg Arg Ser Gly Arg Gln Gln Leu Gln Ser Ile Lys Gly
    130                 135                 140

Thr Lys Leu Thr Ile Thr Gln Ala Val Thr Thr Thr Thr Thr Trp Arg
145                 150                 155                 160

Pro Ser Ser Thr Thr Thr Ile Ala Gly Leu Arg Val Thr Glu Ser Lys
                165                 170                 175

Gly His Ser Glu Ser Trp His Leu Ser Leu Asp Thr Ala Ile Arg Val
            180                 185                 190

Ala Leu Ala Val Ala Val Leu Lys Thr Val Ile Leu Gly Leu Leu Cys
        195                 200                 205

Leu Leu Leu Trp Trp Arg Arg Arg Lys Gly Ser Arg Ala Pro Ser Ser
    210                 215                 220

Asp Phe
225
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 678 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..678

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATG GGT CGG CCC CTG CTG CTG CCC CTG CTG CTC CTG CTG CAG CCG CCA      48
Met Gly Arg Pro Leu Leu Leu Pro Leu Leu Leu Leu Leu Gln Pro Pro
  1               5                  10                  15
```

```
GCA TTT CTG CAG CCT GGT GGC TCC ACA GGA TCT GGT CCA AGC TAC CTT          96
Ala Phe Leu Gln Pro Gly Gly Ser Thr Gly Ser Gly Pro Ser Tyr Leu
             20                  25                  30

TAT GGG GTC ACT CAA CCA AAA CAC CTC TCA GCC TCC ATG GGT GGC TCT         144
Tyr Gly Val Thr Gln Pro Lys His Leu Ser Ala Ser Met Gly Gly Ser
         35                  40                  45

GTG GAA ATC CCC TTC TCC TTC TAT TAC CCC TGG GAG TTA GCC ACA GCT         192
Val Glu Ile Pro Phe Ser Phe Tyr Tyr Pro Trp Glu Leu Ala Thr Ala
     50                  55                  60

CCC GAC GTG AGA ATA TCC TGG AGA CGG GGC CAC TTC CAC GGG CAG TCC         240
Pro Asp Val Arg Ile Ser Trp Arg Arg Gly His Phe His Gly Gln Ser
 65                  70                  75                  80

TTC TAC AGC ACA AGG CCG CCT TCC ATT CAC AAG GAT TAT GTG AAC CGG         288
Phe Tyr Ser Thr Arg Pro Pro Ser Ile His Lys Asp Tyr Val Asn Arg
                 85                  90                  95

CTC TTT CTG AAC TGG ACA GAG GGT CAG GAG AGC GGC TTC CTC AGG ATC         336
Leu Phe Leu Asn Trp Thr Glu Gly Gln Glu Ser Gly Phe Leu Arg Ile
            100                 105                 110

TCA AAC CTG CGG AAG GAG GAC CAG TCT GTG TAT TTC TGC CGA GTC GAG         384
Ser Asn Leu Arg Lys Glu Asp Gln Ser Val Tyr Phe Cys Arg Val Glu
        115                 120                 125

CTG GAC ACC CGG AGA TCA GGG AGG CAG CAG TTG CAG TCC ATC AAG GGG         432
Leu Asp Thr Arg Arg Ser Gly Arg Gln Gln Leu Gln Ser Ile Lys Gly
    130                 135                 140

ACC AAA CTC ACC ATC ACC CAG GCT GTC ACA ACC ACC ACC ACC TGG AGG         480
Thr Lys Leu Thr Ile Thr Gln Ala Val Thr Thr Thr Thr Thr Trp Arg
145                 150                 155                 160

CCC AGC AGC ACA ACC ACC ATA GCC GGC CTC AGG GTC ACA GAA AGC AAA         528
Pro Ser Ser Thr Thr Thr Ile Ala Gly Leu Arg Val Thr Glu Ser Lys
                165                 170                 175

GGG CAC TCA GAA TCA TGG CAC CTA AGT CTG GAC ACT GCC ATC AGG GTT         576
Gly His Ser Glu Ser Trp His Leu Ser Leu Asp Thr Ala Ile Arg Val
            180                 185                 190

GCA TTG GCT GTC GCT GTG CTC AAA ACT GTC ATT TTG GGA CTG CTG TGC         624
Ala Leu Ala Val Ala Val Leu Lys Thr Val Ile Leu Gly Leu Leu Cys
        195                 200                 205


CTC CTC CTG TGG TGG AGG AGA AGG AAA GGT AGC AGG GCG CCA AGC AGT         672
Leu Leu Leu Trp Trp Arg Arg Arg Lys Gly Ser Arg Ala Pro Ser Ser
    210                 215                 220
```

```
GAC TTC                                                                678
Asp Phe
225
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 933 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 55..813

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
TCGACGCGCC TATTCTTGCA CGAGGCAGAC AGGCTGTGCC AGAACTCTTC AGAG ATG      57
                                                           Met
                                                             1

TCA TTT TCT TTC TCT CAA CCA TGC CCC TCA GGG GCA CTT CTG CTG GTG     105
Ser Phe Ser Phe Ser Gln Pro Cys Pro Ser Gly Ala Leu Leu Leu Val
               5                  10                  15

GTG GTG TCA AGC CTC CTT TTA TGG GAG AAT GTG GCC TCT GTA CCT TTG     153
Val Val Ser Ser Leu Leu Leu Trp Glu Asn Val Ala Ser Val Pro Leu
         20                  25                  30

AGT AGC AAT GAG ACT GAT GGT TAT CCA TTA TCC ATC AAT GGG CTG TTT     201
Ser Ser Asn Glu Thr Asp Gly Tyr Pro Leu Ser Ile Asn Gly Leu Phe
     35                  40                  45

CAT AAT GCC ATG AGA CTA ACT TGG AAT ATC AAA AAC CTC AAC ATG GAA     249
His Asn Ala Met Arg Leu Thr Trp Asn Ile Lys Asn Leu Asn Met Glu
 50                  55                  60                  65

CTG CGC AAG ACA TAT ACA GTC AAT CAA GTC TCT GAA AAA TTA TAC GAG     297
Leu Arg Lys Thr Tyr Thr Val Asn Gln Val Ser Glu Lys Leu Tyr Glu
                 70                  75                  80

AAC TAT ATG CTT GAC TTT ATT GAG GAC ATG GAG TAT CTG GTC AAG GCT     345
Asn Tyr Met Leu Asp Phe Ile Glu Asp Met Glu Tyr Leu Val Lys Ala
             85                  90                  95

CTC ACC TGC TGC CAC AAT TAT TCC ATC AAA ACT CCA GAA AAC CTG GAC     393
Leu Thr Cys Cys His Asn Tyr Ser Ile Lys Thr Pro Glu Asn Leu Asp
        100                 105                 110

GAA GCT CAA CAG ATT CCT TTT AAC GAA TTT CCA AAG CTG ATC CTC AGT     441
Glu Ala Gln Gln Ile Pro Phe Asn Glu Phe Pro Lys Leu Ile Leu Ser
    115                 120                 125
```

```
AGA ATG TGG GCT TGG AAT GAA ACT TCT AAA GTT CTA CTG ACC ACA CTC      489
Arg Met Trp Ala Trp Asn Glu Thr Ser Lys Val Leu Leu Thr Thr Leu
130                 135                 140                 145

AGA AGT ATT CCA GGA ATG CAT GAT GAT GTC ATT TCA TTA GCC AAA AAC      537
Arg Ser Ile Pro Gly Met His Asp Asp Val Ile Ser Leu Ala Lys Asn
                150                 155                 160

ATT GAA ACA AAA CTT GCA GAG CTT TTT GAG TAC ACC CAG AGT ATA CTC      585
Ile Glu Thr Lys Leu Ala Glu Leu Phe Glu Tyr Thr Gln Ser Ile Leu
            165                 170                 175

AAC TCG ATT TAT GGA ACA ACA ACA ACA GGA AAT GTG GAA TAC ACC GTC      633
Asn Ser Ile Tyr Gly Thr Thr Thr Thr Gly Asn Val Glu Tyr Thr Val
        180                 185                 190

TTT TCT GGT CTT GAA GAC TTA AAA TCA TCT GAT GAA GAA TTT AGT CTT      681
Phe Ser Gly Leu Glu Asp Leu Lys Ser Ser Asp Glu Glu Phe Ser Leu
    195                 200                 205

TTT GAC CTT TGT AAA TTT TCC TAT TGC TTA CGT GTA GAT ATA CAT ATG      729
Phe Asp Leu Cys Lys Phe Ser Tyr Cys Leu Arg Val Asp Ile His Met
210                 215                 220                 225

GTT GAA CTT TAT CTC AAG CTA TTA GAG TGT GTG GTA TAT GTT AGT AGT      777
Val Glu Leu Tyr Leu Lys Leu Leu Glu Cys Val Val Tyr Val Ser Ser
                230                 235                 240

GAT GTT TGT TTA TCC AAA AAT ATT AGA GAT GCT TCA TGATGCTGAA           823
Asp Val Cys Leu Ser Lys Asn Ile Arg Asp Ala Ser
            245                 250

TCTTTTTAAA TAATCTTAAT TTTATAATTG TGAAAGTATA ATTGAGTATA ACGAGTGTCT   883

TTTAAAATAA AAATAAACTA TATATATAAA AAAAAAAAAA AAAAAAAAAA              933
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 253 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Met Ser Phe Ser Phe Ser Gln Pro Cys Pro Ser Gly Ala Leu Leu Leu
  1               5                  10                  15

Val Val Val Ser Ser Leu Leu Leu Trp Glu Asn Val Ala Ser Val Pro
            20                  25                  30
```

```
Leu Ser Ser Asn Glu Thr Asp Gly Tyr Pro Leu Ser Ile Asn Gly Leu
          35                  40                  45

Phe His Asn Ala Met Arg Leu Thr Trp Asn Ile Lys Asn Leu Asn Met
      50                  55                  60

Glu Leu Arg Lys Thr Tyr Thr Val Asn Gln Val Ser Glu Lys Leu Tyr
 65                  70                  75                  80

Glu Asn Tyr Met Leu Asp Phe Ile Glu Asp Met Glu Tyr Leu Val Lys
                 85                  90                  95

Ala Leu Thr Cys Cys His Asn Tyr Ser Ile Lys Thr Pro Glu Asn Leu
            100                 105                 110

Asp Glu Ala Gln Gln Ile Pro Phe Asn Glu Phe Pro Lys Leu Ile Leu
        115                 120                 125

Ser Arg Met Trp Ala Trp Asn Glu Thr Ser Lys Val Leu Leu Thr Thr
    130                 135                 140

Leu Arg Ser Ile Pro Gly Met His Asp Asp Val Ile Ser Leu Ala Lys
145                 150                 155                 160

Asn Ile Glu Thr Lys Leu Ala Glu Leu Phe Glu Tyr Thr Gln Ser Ile
                165                 170                 175

Leu Asn Ser Ile Tyr Gly Thr Thr Thr Thr Gly Asn Val Glu Tyr Thr
            180                 185                 190

Val Phe Ser Gly Leu Glu Asp Leu Lys Ser Ser Asp Glu Glu Phe Ser
        195                 200                 205

Leu Phe Asp Leu Cys Lys Phe Ser Tyr Cys Leu Arg Val Asp Ile His
    210                 215                 220

Met Val Glu Leu Tyr Leu Lys Leu Leu Glu Cys Val Val Tyr Val Ser
225                 230                 235                 240

Ser Asp Val Cys Leu Ser Lys Asn Ile Arg Asp Ala Ser
                245                 250
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 762 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..759

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
ATG TCA TTT TCT TTC TCT CAA CCA TGC CCC TCA GGG GCA CTT CTG CTG      48
Met Ser Phe Ser Phe Ser Gln Pro Cys Pro Ser Gly Ala Leu Leu Leu
  1               5                  10                  15

GTG GTG GTG TCA AGC CTC CTT TTA TGG GAG AAT GTG GCC TCT GTA CCT      96
Val Val Val Ser Ser Leu Leu Leu Trp Glu Asn Val Ala Ser Val Pro
            20                  25                  30

TTG AGT AGC AAT GAG ACT GAT GGT TAT CCA TTA TCC ATC AAT GGG CTG     144
Leu Ser Ser Asn Glu Thr Asp Gly Tyr Pro Leu Ser Ile Asn Gly Leu
        35                  40                  45

TTT CAT AAT GCC ATG AGA CTA ACT TGG AAT ATC AAA AAC CTC AAC ATG     192
Phe His Asn Ala Met Arg Leu Thr Trp Asn Ile Lys Asn Leu Asn Met
    50                  55                  60

GAA CTG CGC AAG ACA TAT ACA GTC AAT CAA GTC TCT GAA AAA TTA TAC     240
Glu Leu Arg Lys Thr Tyr Thr Val Asn Gln Val Ser Glu Lys Leu Tyr
65                  70                  75                  80

GAG AAC TAT ATG CTT GAC TTT ATT GAG GAC ATG GAG TAT CTG GTC AAG     288
Glu Asn Tyr Met Leu Asp Phe Ile Glu Asp Met Glu Tyr Leu Val Lys
                85                  90                  95

GCT CTC ACC TGC TGC CAC AAT TAT TCC ATC AAA ACT CCA GAA AAC CTG     336
Ala Leu Thr Cys Cys His Asn Tyr Ser Ile Lys Thr Pro Glu Asn Leu
            100                 105                 110

GAC GAA GCT CAA CAG ATT CCT TTT AAC GAA TTT CCA AAG CTG ATC CTC     384
Asp Glu Ala Gln Gln Ile Pro Phe Asn Glu Phe Pro Lys Leu Ile Leu
        115                 120                 125

AGT AGA ATG TGG GCT TGG AAT GAA ACT TCT AAA GTT CTA CTG ACC ACA     432
Ser Arg Met Trp Ala Trp Asn Glu Thr Ser Lys Val Leu Leu Thr Thr
    130                 135                 140

CTC AGA AGT ATT CCA GGA ATG CAT GAT GAT GTC ATT TCA TTA GCC AAA     480
Leu Arg Ser Ile Pro Gly Met His Asp Asp Val Ile Ser Leu Ala Lys
145                 150                 155                 160

AAC ATT GAA ACA AAA CTT GCA GAG CTT TTT GAG TAC ACC CAG AGT ATA     528
Asn Ile Glu Thr Lys Leu Ala Glu Leu Phe Glu Tyr Thr Gln Ser Ile
                165                 170                 175

CTC AAC TCG ATT TAT GGA ACA ACA ACA ACA GGA AAT GTG GAA TAC ACC     576
Leu Asn Ser Ile Tyr Gly Thr Thr Thr Thr Gly Asn Val Glu Tyr Thr
            180                 185                 190

GTC TTT TCT GGT CTT GAA GAC TTA AAA TCA TCT GAT GAA GAA TTT AGT     624
Val Phe Ser Gly Leu Glu Asp Leu Lys Ser Ser Asp Glu Glu Phe Ser
        195                 200                 205
```

```
CTT TTT GAC CTT TGT AAA TTT TCC TAT TGC TTA CGT GTA GAT ATA CAT      672
Leu Phe Asp Leu Cys Lys Phe Ser Tyr Cys Leu Arg Val Asp Ile His
    210                 215                 220

ATG GTT GAA CTT TAT CTC AAG CTA TTA GAG TGT GTG GTA TAT GTT AGT        7
Met Val Glu Leu Tyr Leu Lys Leu Leu Glu Cys Val Val Tyr Val Ser
225                 230                 235                 240

AGT GAT GTT TGT TTA TCC AAA AAT ATT AGA GAT GCT TCA                  759
Ser Asp Val Cys Leu Ser Lys Asn Ile Arg Asp Ala Ser
                245                 250
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 523 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..261

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
GGA TCG ACC CCC GCG TCC GCG GTT TCT GCG CGG TGC GCG ACC TCG TCC       48
Gly Ser Thr Pro Ala Ser Ala Val Ser Ala Arg Cys Ala Thr Ser Ser
  1               5                  10                  15

CGA AGC CTG GGG ATA CAC CCT CTC GAG AGC CCG CTG TCG CCC TCC GTT       96
Arg Ser Leu Gly Ile His Pro Leu Glu Ser Pro Leu Ser Pro Ser Val
            20                  25                  30

AAG GTC GAA CCC CTC ACA GTT GCT GTG GGC AAC TCC AGC CCA ACA TTC      144
Lys Val Glu Pro Leu Thr Val Ala Val Gly Asn Ser Ser Pro Thr Phe
        35                  40                  45

CCT CGC TCT GGT TCT CGC CCC ATT GGG AAA CTC GGC CCC ACG CTT CCC      192
Pro Arg Ser Gly Ser Arg Pro Ile Gly Lys Leu Gly Pro Thr Leu Pro
    50                  55                  60


ACT TTT CTG GAT GAG GTG TCC CCT TTC TCC CCA CTA AAA TGT CAA ATA      240
Thr Phe Leu Asp Glu Val Ser Pro Phe Ser Pro Leu Lys Cys Gln Ile
65                  70                  75                  80

ACC TAC GGA GGG TCT TCC TGAAACCCGC AGAGGAAAAT TCAGGCAACG             288
Thr Tyr Gly Gly Ser Ser
                85

CCTCGCGTTG TGGTTTCAGG CTGCATGTAC CAAGTAGTTC AGACGTTTGG CTCGGATGGA   348
```

AAAAAATCTT CTGCAATTAT TTCCAATTCC TAAGTCTTCT TGGAAATCTT ATACCACTAG 408

TTCAATCTTC AAGTCATGTC TGATGCTTTG AAAGGGATTA CAGGAAACCA GTTCAAGTTA 468

TTTTCAGACC AGATTTCCAG CTCTTCCACA AGTGCATCAT TTCAATTGCC CATTT 523

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gly Ser Thr Pro Ala Ser Ala Val Ser Ala Arg Cys Ala Thr Ser Ser
1 5 10 15

Arg Ser Leu Gly Ile His Pro Leu Glu Ser Pro Leu Ser Pro Ser Val
20 25 30

Lys Val Glu Pro Leu Thr Val Ala Val Gly Asn Ser Ser Pro Thr Phe
35 40 45

Pro Arg Ser Gly Ser Arg Pro Ile Gly Lys Leu Gly Pro Thr Leu Pro
50 55 60

Thr Phe Leu Asp Glu Val Ser Pro Phe Ser Pro Leu Lys Cys Gln Ile
65 70 75 80

Thr Tyr Gly Gly Ser Ser
85

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 258 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGATCGACCC CCGCGTCCGC GGTTTCTGCG CGGTGCGCGA CCTCGTCCCG AAGCCTGGGG 60

ATACACCCTC TCGAGAGCCC GCTGTCGCCC TCCGTTAAGG TCGAACCCCT CACAGTTGCT 120

GTGGGCAACT CCAGCCCAAC ATTCCCTCGC TCTGGTTCTC GCCCCATTGG GAAACTCGGC 180

CCCACGCTTC CCACTTTTCT GGATGAGGTG TCCCCTTTCT CCCCACTAAA ATGTCAAATA 240

ACCTACGGAG GGTCTTCC 258

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 243 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Leu Pro Ser Leu Ile Gln Pro Cys Ser Ser Gly Thr Leu Leu Met
1 5 10 15

Leu Leu Met Ser Asn Leu Phe Leu Trp Glu Lys Val Ser Ser Ala Pro
20 25 30

Ile Asn Ala Ser Glu Ala Val Leu Ser Asp Leu Lys Asp Leu Phe Asp
35 40 45

Asn Ala Thr Val Leu Ser Gly Glu Met Ser Lys Leu Gly Val Ile Met
50 55 60

Arg Lys Glu Phe Phe Met Asn Ser Phe Ser Ser Glu Thr Phe Asn Lys
65 70 75 80

Ile Ile Leu Asp Leu His Lys Ser Thr Glu Asn Ile Thr Lys Ala Phe
85 90 95

Asn Ser Cys His Thr Val Pro Ile Asn Val Pro Glu Thr Val Glu Asp
100 105 110

Val Arg Lys Thr Ser Phe Glu Glu Phe Leu Lys Met Val Leu His Met
115 120 125

Leu Leu Ala Trp Lys Glu Pro Leu Lys His Leu Val Thr Glu Leu Ser
130 135 140

Ala Leu Pro Glu Cys Pro Tyr Arg Ile Leu Ser Lys Ala Glu Ala Ile
145 150 155 160

Glu Ala Lys Asn Lys Asp Leu Leu Glu Tyr Ile Ile Arg Ile Ile Ser
165 170 175

Lys Val Asn Pro Ala Ile Lys Glu Asn Glu Asp Tyr Pro Thr Trp Ser
180 185 190

Asp Leu Asp Ser Leu Lys Ser Ala Asp Lys Glu Thr Gln Phe Phe Ala
195 200 205

Leu Tyr Met Phe Ser Phe Cys Leu Arg Ile Asp Leu Glu Thr Val Asp
210 215 220

Phe Leu Val Asn Phe Leu Lys Cys Leu Leu Leu Tyr Asp Asp Val Cys
225 230 235 240

Tyr Ser Glu

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1-18
(D) OTHER INFORMATION: /note= "forward PCR primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TCACTCAACC AAAACACC 18

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1-15
(D) OTHER INFORMATION: /note= "reverse PCR primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CAGTTCAGAA AGACC 15

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/21151

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :C12Q 21/06, 1/68; C12N 15/00, 1/12, 5/00, 5/02
US CL : 435/69.1, 320.1, 252.1, 6, 325
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 320.1, 252.1, 6, 325

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, HCAPLUS, BIOSIS, EMBASE, SCISEARCH, NTIS, WPIDS, LIFESCI, BIOTECHDS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	LIN et al. Two novel members of the Prolactin/Growth hormone Family are expressed in the mouse placenta. Endocrinology. 21 November 1997, Vol. 138, pages 5535-5540, see Figures 1-3.	1-15, 20-22
X	ADAMS et al. Initial assesment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. 28 September 1995, Vol. 377 (6547 suppl), pages 3-17, see the entire article, specially EST 64078.	5-15, 20-22
X,P	Database Genbank on STN, Acession No. AF20524, Mus musculus prolactin-like protein F precursor. 09 October 1997, see the entire citation.	5-15, 20-22

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
28 JANUARY 1999

Date of mailing of the international search report
24 FEB 1999

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
MARYAM MONSHIPOUR
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)★

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/21151

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-15, 20-22

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/21151

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

Group I, claims 1-15 and 20-22, drawn to an isolated nucleic acid encoding a human LSP-1 poltpeptide, a method of detecting the nucleic acids using the hybridization assay, a method of expressing the polypeptide and the isolated human LSP-1 polypeptide.

Group II, claims 16-19, drawn to an antibody rerived against the polypeptide, and a method of detecting the polypeptide using the antibody.

Group III, claims 23-24, drawn to a method of identifying ligands which bind the polypeptide.

Group IV, claims 25-26, drawn to a method of modulating the activity of the polypeptide and a method of identifying the modulatore of the polypeptide.

Groups I-IV do not relate to a single inventive concept because they are considered to be four different categories of invention and are not drawn to combinations of categories (i.e. categories 1-5), specified in 37 CFR section 1.475(b).